A Robust High-dimensional Data Reduction Method

Longcun Jin\(^1\), Wanggen Wan\(^1\), Yongliang Wu\(^1\), Bin Cui\(^1\) Xiaoqing Yu\(^1\) and Youyong Wu\(^{1,2}\)

\(^1\)School of Communication and Information Engineering, Shanghai University, Shanghai, China
\(^2\)School of Information and Electronic Engineering, East China University of Technology, Fuzhou, Jiangxi, China

Abstract—In this paper, we propose a robust high-dimensional data reduction method. The model assumes that the pixel reflectance results from linear combinations of pure component spectra contaminated by an additive noise. The abundance parameters appearing in this model satisfy positivity and additive constraints. These constraints are naturally expressed in a Bayesian literature by using appropriate abundance prior distributions. The posterior distributions of the unknown model parameters are then derived. The proposed algorithm consists of Bayesian inductive cognition part and hierarchical reduction algorithm model part. The proposed reduction algorithm based on Bayesian inductive cognitive model is used to decide which dimensions are advantageous and to output the recommended dimensions of the hyperspectral image. The algorithm can be interpreted as a robust reduction inference method for a Bayesian inductive cognitive model. Experimental results on high-dimensional data demonstrate useful properties of the proposed reduction algorithm.

Index Terms—High-dimensional data, Bayesian model, Inductive cognitive, Hyperspectral image.

I. INTRODUCTION

The reduction of hyperspectral image has been widely used in remote sensing signal processing for data analysis [1-4]. Its underlying assumption is based on the fact that all data sample vectors are mixed by a number of so-called endmembers assumed to be presented in the data. By virtue of this assumption, two models have been investigated in the past to model how mixing activities take place. One is the macrospectral mixture that describes a mixed pixel as a linear mixture endmembers opposed to the other model suggested by Hapke, referred to as intimate mixture that models a mixed pixel as a nonlinear mixture [5-6]. Another is how to estimate these endmembers once the number of endmembers is determined. Such statistics are said to be sufficient if they capture all the “relevant information” in the sample about the identity of reduction of hyperspectral image [7-8].

In this work we propose a way of quantifying this information using information theoretic notions and show how features which maximize this information can be extracted. Due to its link with the statistical concept of reduction, an view of this setup is a generalization of the problem of nonlinear regression. There is often more information between \(X\) and \(Y\) than can be captured by a single conditional expectation value [9-13]. Our problem then is to find several such functions, or regressions, that together capture more of the information and structure of the variables. As will be shown in this paper, this problem can be cast as finding dimension reduction of hyperspectral image which captures the mutual information in a two-way contingency table. It is thus related to a long line of work in statistics.

As explained above, the linear reduction of model is classically used to model the spectrum of a pixel in the observed scene. And the dimension reduction of hyperspectral image has already received much attention in the literature [3-4, 14-17]. Consequently, their estimation requires to use a quadratic programming algorithm with linear equalities and inequalities as constraints. Different estimators were developed using these ideas [18-20]. This paper studies a Bayesian estimator of Inductive Cognition which allows one to estimate the abundances in an nonlinear reduction of model. The proposed algorithm defines appropriate prior distributions for the unknown signal parameters and estimates these unknown parameters from their posterior distributions.

The prior distributions used in the present paper depend on hyperparameters which have to be determined. The unknown parameters in the posterior distributions requires to use appropriate simulation methods such as hierarchical Bayesian methods [21-23]. It is well known that the dimensionality of input space strongly affects performance of many classification methods. This requires the careful design of new algorithms that are able to handle hundreds of such spectral images at the same time minimizing the effects from the “curse of dimensionality”. We first illustrate a well-known phenomenon in hyperspectral data. To address the characteristic that certain parts of the spectrum will provide a much richer descriptor for classification than other parts, some approaches such as a straightforward feature selection or a block-based approximation to the covariance matrix can be applied [24-25].

The remainder of the paper is organized as follows. An overview of the related work is given in Section II. The proposed algorithm is discussed in Section III. Section IV presents the experimental settings and performance evaluation. Section V concludes this paper.

II. RELATED WORK

The natural variability of the material spectra and the noise added by the transmission media and sensor system make necessary the use of statistical methods for information extraction.
and pattern recognition on hyperspectral data. Hyperspectral imaging technology has found application beyond earth remote sensing in agriculture, medicine, biology, pharmaceuticals, forensics, color vision, target detection, archaeology, and many others near field applications. However, classification of hyperspectral data is primarily made on a pixel by pixel basis with classification accuracy figures in the range 79% - 84%, and they have not changed significantly in recent decade [4, 6, 10, 26-28]. The scale-space framework introduced by the diffusion equation has been also used for image reduction, in conjunction with level sets to detect movement in image sequences, information extraction and image restoration, registration, and classification integrating level sets in a common framework [17-18]. Continuous transformation of the original image into a space of progressively smoother images identified by the scale or level of image smoothing, in terms of pixel resolution.

In recent years, hyperspectral data reduction and classification are generally considered as a processing step for spectro classification, target detection or segmentation. For each of these problems, different methods exist. Methods of feature extraction like projection pursuit or Bayesian construct a lower dimension feature space by transforming the original data to preserve the most discriminable information content but change the physical meaning of the components [10, 13, 15, 20]. Spectral reduction methods, as discussed in [4, 13], reduce the number of spectra by keeping the representative ones. Such methods do not care about the spatial structure of hyperspectral data; hyperspectral data is only considered as a set of spectra. In [3], the authors use independent component analysis for band selection reduction method. The aim in this case is to identify the spectral signature that identifies these materials, each one having specific characteristics on certain bands. [24] utilized a Gaussian mixture model to analysis to reduce the size of the spectral vector. In our study, we account for the spatial structure of the data through a hierarchical Bayesian model about from that we also account for the noise and thus estimate a typical mean spectra for each class. The authors in [7] show that the reduction matrix obtained might be far from the true one. The classical hierarchical Bayesian model does not take into account any information about noise or any prior information on the signal of interest [29-31]. In this paper, we proposed a Bayesian estimation framework for the sources with a common Bayesian classification variable which is modeled as an inductive cognition field. Each value of this variable corresponds to a characteristic mean spectrum of a given region.

III. THE PROPOSED ALGORITHM

Since Bayesian model with concentration hyperparameter $\alpha$ defines a prior on all partitions of the $n_i$ hyperspectral image point $D$ (the value of $\alpha$ is directly related to the expected number of classes), the prior on the merged hypothesis is the relative mass of all $n_i$ points belonging to one class versus all the other partitions of those $n_i$ virtual reality data points consistent with the tree structure. The proposed algorithm is based Bayesian inductive cognitive model by an approximate inference approach. This can be computed bottom-up as the tree built shown in Fig. 1.

**Algorithm**

1. Initialize each leaf $i$ to have $d_i = \alpha \cdot \pi_i = 1$.
2. For each internal node $k$ do:
   - $d_k = \alpha \Gamma(n_k) + d_{left} d_{right}$
   - $\pi_k = \frac{\alpha \Gamma(n_k)}{d_k}$
3. End for

**Fig. 1.** To compute prior on merging, where $right$ ($left$) indexes the right (left) subtree of $T_i$ and $d_{left}, (d_{right})$ is the value of $d$ computed for the right (left) child of internal nodes $k$.

The main idea in our Bayesian of inductive cognition model is to use the hyperspectral model (3) and the prior distributions (4) and

\[
\pi(\Lambda) = N(\Lambda, \Sigma) \\
\pi(\Sigma_{i} | \nu_{i}, \Sigma_{i}) = |\Sigma_{i}|^{-\frac{\nu_{i}}{2}} \exp \left( -\frac{1}{2} \text{Tr} (\Sigma_{i}^{-1} \mathbf{x}) \right) \\
\pi(\mu_{i}) = N(\mu_{i}, \nu_{i}) \\
\pi(\sigma^{2}_{i}) = IG(\alpha_{i}, \gamma_{i})
\]

to obtain the posterior law

\[
p(\mathbf{s}, \mathbf{z}, \theta | x) \propto p(\mathbf{x} | \mathbf{s}, \mathbf{z}, \theta) p(\mathbf{s} | \mathbf{z}, \theta) p(\theta)
\]

where $\theta = (\theta_{\Lambda}, \theta_{\Sigma})$ and $\theta = (\Lambda, \Sigma_{i}, \mu_{i}, \sigma^{2}_{i})$ represent the mixing matrix $\Lambda$ and all the hyperparameters of the prior modeling of the noise and the sources. Here, the unknowns are first splitted in two categories: the Bayesian variables $(\mathbf{s}, \mathbf{z})$ and all the hyperparameters $\theta$, and then use the following iterative algorithm. Estimate $(\hat{s}, \hat{z})$ using $p(\mathbf{s}, \mathbf{z} | \theta_{\Lambda}, \theta_{\Sigma}, x)$ by

$$\hat{s} - p(s | \hat{z}, \theta_{\Lambda}, x).$$

In this paper, $\hat{s}$ represents either $\arg \max$ or compute the MFA. To implement this algorithm, we need the following expressions:

$$p(s | z, \theta_{\Lambda}, \theta_{\Sigma}, x) \propto p(s | z, \theta_{\Lambda}, \theta_{\Sigma})$$

It is then easy to see that $p(s | z, \theta_{\Lambda}, x)$ is separable in $r$. $s(r)$ is estimated by its mean value $\hat{s}(r)$

$$p(s | z, \theta_{\Lambda}, x) = \prod p(s(r) | z(r), \theta_{\Lambda}, x(r)) = \prod N(\hat{s}(r), Bb(r))$$

with

$$B(r) = \begin{bmatrix} A' \Sigma_{i} + \Sigma_{\nu_{i}} \end{bmatrix}^{-1}$$

$$\hat{s}(r) = B(r)A' \Sigma_{i} x(r) + \Sigma_{\nu_{i}} m_{i}(r)$$

In this relation, $m_{i}(r)$ is a vector of size $K$ with all elements equal to zero except the $k$th element where $k = z(r)$ and $\Sigma_{i}(r)$ is a diagonal matrix of size $K \times K$ with all elements equal to zero except the $k$th diagonal where $k = z(r)$. Note that the prior law on $z$ is

$$p(z | \beta) \propto \prod p(z(r) | \pi(r), \beta)$$

$$\propto \exp \left[ \sum_{r \in V(r)} \beta z(r) z(r) \right]$$
The expression of the posterior \( P(z|\theta, x) \) is
\[
P(z|\theta, x) \propto \prod_{r \in V(r)} p(z(x)) p(z(r) | x(r), \theta) p(z(r) | z(r), \theta) \prod_{r \in V(r)} p(z(x)), \quad r \in V(r), \theta, \beta
\]

In this paper, our Bayesian of reductive cognitive model consists then in approximation \( p(z|\theta, x) \) by
\[
q(z|\theta, x) = \prod_{r \in V(r)} q(z(r) | x(r), \theta) q(z(r) | z(r), \theta, \beta)
\]
where \( \bar{z}(r) \) represents the mean value of \( z(r) \), and
\[
q(z(r) | x(r), \bar{z}(r), r \in V(r), \theta, \beta) \\
\approx p(x(r) | z(r), \theta) p(z(r) | x(r), \bar{z}(r), r \in V(r), \theta, \beta) \\
\times \exp \left[ \beta \bar{z}(r) z(r) + \ln p(x(r) | z(r), \theta) \right]
\]

In this procedure, the proposed algorithm is to incorporate the idea of multi-scale hyperspectral data reduction into extended Bayesian of inductive cognitive transformations. However, if the reduction patterns do not have regular properties across the hyperspectral data, an adaptive scheme is needed to ensure the wonderful experimental result. In order to extend our algorithm and closing operations to hyperspectral image, let us consider a hyperspectral image \( f \) defined on \( \mathbb{R}^d \). Given a Bayesian model of minimal size, extended opening by reconstruction is defined by
\[
(f \circ B)^\lambda(x, y) = \bigcap_{s \in \mathbb{Z}} [o_s(f \circ B)(f)](x, y)
\]

where
\[
[o_s(f \circ B)](x, y) = \bigcup_{k=0}^1 \text{dist}_B(f \circ B, k)
\]

The elementary term \( [o_s(f \circ B)](x, y) \) is an extended geodesic dilation, defined as the maximum of the elementary dilation of the elementary dilation of \( f \circ B \) using \( B \) at pixel \( (x, y) \) and the value of \( f(x, y) \),
\[
[o_s(f \circ B, k)](x, y) = \bigcap_{k=0}^1 \text{dist}_B(f \circ B, k)(x, y)
\]

The elementary term \( [o_s(f \circ B)](x, y) \) is an extended geodesic erosion, defined as the minimum of the elementary erosion of \( f \circ B \) using \( B \) at pixel \( (x, y) \) and the value of \( f(x, y) \),
\[
[o_e(f \circ B, k)](x, y) = \bigcup_{k=0}^1 \text{dist}_B(f \circ B, k)(x, y)
\]

Using Eqs. (18) and (20), extended morphological Bayesian models are created as follows. Let the vector \( p_\lambda(x, y) \) be the extended opening by reconstruction Bayesian model at the point \( (x, y) \) of the hyperspectral image \( f \) defined by
\[
p_\lambda(x, y) = (f \circ B)^\lambda(x, y), \quad \lambda = 0, 1, \ldots, k
\]

And let \( p_\lambda(x, y) \) be the extended closing by reconstruction Bayesian model at the point \( (x, y) \) of the hyperspectral image \( f \) defined by
\[
p_\lambda(x, y) = (f \circ B)^\lambda(x, y), \quad \lambda = 0, 1, \ldots, k
\]

IV. EXPERIMENTAL RESULTS

In this part, we will present the effectiveness of the dimensional reduction approach based on Bayesian model, over the hyperspectral image data set. However, since the data is separated into various discontinuous intervals, the dimensional reduction approach cannot be applied as is and need to be adapted for the hyperspectral data. Thus, the regular Bayesian reduction algorithm can be applied as it is, on each interval. The decomposed hyperspectral data output is composed of the decomposed output of each considered interval, appended to each other. This mechanism is illustrated in Fig. 4. The hyperspectral data interval decomposition can be done during the pre-processing stage.

Such algorithm is composed of three major phases: pre-processing, processing and post-processing. The
pre-processing initializes the problem variables and orders the intervals from the hyperspectral data file. The processing step determines for each pixel and its interval the maximum number of Bayesian dimensional reduction which leads to a viable correlation. The granularity of considered intervals used during the Bayesian processing can vary. It can be matching the number of intervals of the hyperspectral data, namely 17 intervals. On the other hand, the granularity can be coarser, to match the three major subintervals. This is possible since intervals composing those subintervals are generating contiguous spectrum ranges. However, a special care must be taken to conserve spectral contiguousness inside each major subinterval, i.e. reading intervals 1, 2, 3, 4 in this order, according to Table 1, for first major subinterval.

Table 1 shows the level of Bayesian dimension reduction achieved for different correlation values, as well as the resulting number of bands, per major sub-interval. Each sub-interval is identified by its interval components. The total reduction line gives the overall reduction efficiency. Table 1 is considering only three sub-intervals, concatenating close intervals to produce contiguous ranges in frequency domain. On the other hand, Table 2 is considering each interval independently. It can be noticed that by breaking the spectrum into smaller domains, a different level of reduction can be achieved for each interval.

In our work, we conduct some experiments in hyperspectral data dimensionality reduction in order to demonstrate the feasibility of the proposed algorithm. Our datacube was acquired using an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The original scene with size of 256×256 pixels, was acquired by the AVIRIS sensor, which is a mixed river/city area in south of China, early in the growing season. The scene comprises 220 spectral channels to be used in the experiments. These laboratory spectra, which were convolved in accordance with AVIRIS wavelength specifications, will be used to assess endmember signature purity in this work. At this point, it is important to note that many of the mineral spectra in the USGS library are not from the south of China area. Thus, the best match between hyperspectral image endmember and one in the USGS library spectra does not necessarily mean the best match the endmember. In addition, some minerals do not occur in pure form in the area, specifically at the 20-m spatial resolution of the sensor. For illustrative purposes, Fig. 3 shows high-dimension data expressed by computer graphics and a reduction result of hyperspectral data using our proposed reduction based on Bayesian inductive cognition model. It is useful that the results are reduced by our proposed algorithm.

V. CONCLUSION

This paper studied Bayesian inductive cognitive model for nonlinear hyperspectral reduction. The relationships between the different hyperspectral spectra were naturally expressed in a Bayesian context by the prior distributions adopted for the proposed model and their parameters. The posterior distribution of the unknown parameters related to the proposed model were.
estimated by a Bayesian sampling strategy. These posterior distributions provided estimates of the unknown parameters but also information about their uncertainties such as standard deviations of confidence intervals. The proposed algorithms were developed depending whether the endmembers belonging to the mixture are known or belong to a known area. Simulation results conducted on real images illustrated the performance of the proposed reduction algorithm based on Bayesian inductive cognitive model. The hierarchical reduction methodologies developed in this paper could be modified to handle more complicated models. For instance, it would be interesting to extend the proposed algorithm to reduce hyperspectral data composed of homogeneous regions surrounded by hyperspectral image by introducing dimensions correlation via Bayesian models.

Future work will investigate the use of complex hyperspectral data for the selection of the number of reduced bands and of the correlation threshold, as well as for detecting data anomalies. The dimension reduction technique based on Bayesian model will also be compared to the PCA technique to estimate the efficiency of each with the hyperspectral data.

ACKNOWLEDGEMENT

This research work is supported by the National High Technology Research and Development Program of China (863 program, 2007AA01Z319), the National Natural Science Foundation of China (60873130, 60872115), Shanghai’s International Technology Foundation (075107035), and the Shanghai’s Key Discipline Development Program (J50104).

REFERENCES


[29] Yanni Kouskoulas, Fawwaz T. Ulaby, Leland E. Pierce. The Bayesian hierarchical classifier (BHC) and its application to short vegetation using


Longcun Jin received the B.S. degree in Telecommunication Engineering in 2005 from Hangzhou Dianzi University. He is currently a Ph.D. candidate in School of Communication and Information Engineering at Shanghai University. His research interests include in the field of data mining and virtual reality.

Wanggen Wan received the Ph.D. degree in Electronic Engineering in 1992 from Xidian University. He is currently a professor in School of Communication and Information Engineering at Shanghai University. His research interests include virtual reality, data mining and machine learning. He is a Fellow of IET and a senior member of IEEE.