

A HYPERSPECTRAL IMAGE CLASSIFICATION METHOD USING ISOMAP AND RVM

Chang Hongwei^{1,*}, Wang Tao¹, Fang Hao¹, Su Yanzhao¹

¹ The Xi'an Research Institute of High Technology, Xi'an 710025, China - changhongweiathjj@163.com
1106100204@qq.com, fanghao_502@163.com, 1418525651@qq.com

Commission III, WG III/4

KEY WORDS: Hyperspectral Image Classification, Manifold Learning, Isometric Feature Mapping, Support Vector Machines, Relevance Vector Machines

ABSTRACT:

Classification is one of the most significant applications of hyperspectral image processing and even remote sensing. Though various algorithms have been proposed to implement and improve this application, there are still drawbacks in traditional classification methods. Thus further investigations on some aspects, such as dimension reduction, data mining, and rational use of spatial information, should be developed. In this paper, we used a widely utilized global manifold learning approach, isometric feature mapping (ISOMAP), to address the intrinsic nonlinearities of hyperspectral image for dimension reduction. Considering the impropriety of Euclidean distance in spectral measurement, we applied spectral angle (SA) for substitute when constructed the neighbourhood graph. Then, relevance vector machines (RVM) was introduced to implement classification instead of support vector machines (SVM) for simplicity, generalization and sparsity. Therefore, a probability result could be obtained rather than a less convincing binary result. Moreover, taking into account the spatial information of the hyperspectral image, we employ a spatial vector formed by different classes' ratios around the pixel. At last, we combined the probability results and spatial factors with a criterion to decide the final classification result. To verify the proposed method, we have implemented multiple experiments with standard hyperspectral images compared with some other methods. The results and different evaluation indexes illustrated the effectiveness of our method.

1. INTRODUCTION

Hyperspectral image (HSI) is a kind of extraordinary remote sensing image which is different from infrared image and other traditional image. We can not only obtain the spatial information of HSI as ordinary image, but also the distinct and more important spectral information of each pixel. Moreover, the spectral resolution of HSI is so high that the reflectance of each pixel can be nearly plotted as a smooth and continuous curve, which brings lots of benefits for research and its application, such as target detection, classification and anomaly detection (Wang Yiting, 2017). Because of its distinct spectral characteristic, the reflectance spectra has been taken as a key and decisive factor in the application of HSI and many algorithms only focus on the spectral information. However, this brings two problems: 1) because of so many bands HSI contains, the dimension of hyperspectral data could be very large, and the computation cost will be very huge, especially when optimization exists; 2) although identifying different materials based on their reflectance spectra is a necessary and effective way to do HSI processing applications (C. I. Chang, 2003), spatial information which could be also important and useful for higher accuracy is ignored. For the first issue, researchers find that there are strong correlation among spectral bands, which means redundancy exists in HSI. Thus we can do dimension reduction (DR) with little information loss. On account of the nonlinearity of HSI, manifold learning based on graph embedding framework is an appropriate method to handle this problem (Ma Li, 2010). Manifold is a

kind of curly space with local Euclidean space properties, such as spherical, curved surface. It's more suitable for HSI data structure compared with Euclidean space. ISOMAP (Tenenbaum, 2000), which is a famous global manifold learning algorithm, applies geodesic distance to describe the geometric relations of data points in high-dimensional space instead of Euclidean distance and it keeps the same data points' correlation well in low-dimensional space. After DR, hyperspectral image classification could be proceeded. A famous and widely used classification algorithm is SVM, and it has been proved efficient for HSI (F. Melgani, 2004). While compared with SVM, RVM (Christopher M. Bishop, 2006) could be potentially better. The advantages of the RVM over the SVM are probabilistic predictions, automatic estimations of parameters, and the possibility of choosing arbitrary kernel functions (M. E. Tipping, 2001). For reasons above, we adopt RVM for hyperspectral classification to obtain precise probabilistic outputs rather than binary results. Then, spatial factor is taken into consideration. We utilize a spatial weight vector to revise the RVM results so as to the results could be more fit for realistic scenery.

The structure of this paper is organized as follows. In the second section, the basic theory of ISOMAP and RVM are described, and taking advantage of spatial information is also introduced. Experiments and analysis are given in section 3, followed by the conclusion in section 4.

* Corresponding author

2. METHODOLOGY

2.1 Isometric Feature Mapping

By assuming that the distribution of the dataset has a low-dimensional embedded manifold structure, the main idea of ISOMAP is to find a representation of the dataset in the low-dimensional space by keeping the geodesic distance among the data points. The calculation of geodesic distance is achieved through neighbourhood graph. So the steps of ISOMAP are as follows:

Step1: Construct neighbourhood graph G . According to the similarity measurement, generally Euclidean distance, of data points, find k nearest neighbourhood points for each point x_i , defined as $X_i = [x_{i1}, \dots, x_{ik}]$.

Step2: Calculate the geodesic distance $D_G = \{d_G(x_i, x_j), i, j = 1, \dots, N\}$. First initialize the geodesic distance D_G : in neighbourhood graph G , for $x_j \in X_i$, set $d_G(x_i, x_j) = d_E(x_i, x_j)$. Where $d_E(x_i, x_j)$ stands for the Euclidean distance between x_i and x_j . While for $x_j \notin X_i$, we use the shortest path distance calculated by Dijkstra method as the geodesic distance $d_G(x_i, x_j)$.

Step3: Construct low-dimensional embedding Y . Apply Multi-Dimensional Scaling (MDS) method with D_G to minimize the difference of data points before and after embedding

$$\min J = \|\tau(D_G) - \tau(D_{E,Y})\|^2 \quad (1)$$

Where $D_{E,Y}$ is the distance matrix of data points after DR, and

$$\tau(D_G) = -HD_G^2H / 2 \quad (2)$$

$$H = I - ee^T / N \quad (3)$$

Where e is a column vector which all elements are 1.

Manipulate eigenvalue decomposition of $\tau(D_G)$, and select d largest eigenvalues and corresponding eigenvectors to form a diagonal eigenvalue matrix A_d and an eigenvector matrix V_d . Then, we get the final low-dimensional data

$$Y = A_d^{1/2}V_d \quad (4)$$

Though ISOMAP use geodesic distance to describe the correlation of HSI data points, the identity of its measurements is also Euclidean distance. While for HSI, SA is a similarity measurement of spectral waveform. It has good resistance to multiplicative interference and can weaken the influence of light intensity (Freek van der Meer, 2005). Therefore we can just take SA for institute when calculate neighbourhood graph and geodesic distance.

2.2 Relevance Vector Machines

The RVM introduces a prior over the model weights governed by a set of hyperparameters, in a probabilistic framework. One hyperparameter is associated with each weight, and the most probable values are iteratively estimated from the training data. The most compelling feature of the RVM is that it typically

utilizes significantly fewer kernel functions compared to the SVM, while providing a similar performance (Begim Demir, 2007; Psorakis I, 2010).

Taking binary classification for example, assume a training data set $\{x_n, t_n\}_{n=1}^N$, where $\{x_n\}_{n=1}^N$ are the sample vectors, $\{t_n\}_{n=1}^N$ are the corresponding target labels, $t_n \in \{0, 1\}$, and N is the number of training samples (note that the symbols in this subsection are different from those in former subsection). The classification model function of RVM is

$$y(x; w) = \sum_{i=1}^N w_i K(x, x_i) + w_0 \quad (5)$$

Where $K(\cdot; \cdot)$ is kernel function, $w = (w_0, w_1, \dots, w_N)^T$ is the weight vector. Besides, logistics sigmoid link function $\sigma(y) = 1/(1+e^{-y})$ is applied to $y(x; w)$ to obtain probabilistic outputs.

A Bernoulli distribution can be adopted for $p(t|x)$ in the probabilistic framework because only two values (0 and 1) are possible. The likelihood estimation function can be defined as

$$p(t|w) = \prod_{n=1}^N \sigma\{y(x_n; w)\}^{t_n} [1 - \sigma\{y(x_n; w)\}]^{1-t_n} \quad (6)$$

The likelihood is complemented by a prior over the parameters (weights) in the form of

$$p(w|\alpha) = \prod_{n=1}^N \frac{\sqrt{\alpha_n}}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_n w_n^2}{2}\right) \quad (7)$$

Where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$ shows the hyperparameters introduced to control the strength of the prior over its associated weight. Hence, the prior is Gaussian, but conditioned on α . For a certain α value, the posterior weight distribution conditioned on the data is

$$p(w|t, \alpha) = \frac{p(t|w)p(w|\alpha)}{p(t|\alpha)} \quad (8)$$

Where $p(t|w)$ is the likelihood, $p(w|\alpha)$ is the prior, and $p(t|\alpha)$ is referred to as evidence.

As for the weights that couldn't be simply calculated, we need to use a Laplacian approximation procedure. We first define the parameter α , so the most probable weights w_{MP} can be obtained by maximizing $p(w|t, \alpha)$. Because $p(w|t, \alpha)$ is linearly proportional to $p(t|w) \times p(w|\alpha)$, w_{MP} can be calculated by

$$\begin{aligned} \log p(w|t, \alpha) &= \log \{p(t|w)p(w|\alpha)\} - \log p(t|\alpha) \\ &= \sum_{n=1}^N [t_n \log y_n + (1-t_n) \log (1-y_n)] - \frac{1}{2} w^T A w \end{aligned} \quad (9)$$

Where $y_n = \sigma\{y(x_n; w)\}$ and $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ being composed of the current values of α . The iteratively reweighted least-squares algorithm (Christopher M. Bishop, 2006) can be used to find w_{MP} . For this algorithm, we need to obtain the Hessian matrix

$$\nabla_w \nabla_w \log p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha})|_{MP} = -(\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A}) \quad (10)$$

where $\mathbf{B} = \text{diag}(\beta_0, \beta_1, \dots, \beta_N)$

$$\beta_n = \sigma\{y(\mathbf{x}_n)\} [1 - \sigma\{y(\mathbf{x}_n)\}]$$

$$\boldsymbol{\Phi}_{nm} = K(\mathbf{x}_n, \mathbf{x}_{m-1})$$

$$\boldsymbol{\Phi}_{n1} = 1$$

This result is used for a Gaussian approximation to the posterior over weights centered at \mathbf{w}_{MP} :

$$\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A})^{-1} \quad (11)$$

In this way, the classification problem is locally linearized around \mathbf{w}_{MP} in an effective way with

$$\mathbf{w}_{MP} = \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{B} \hat{\mathbf{t}} \quad (12)$$

$$\hat{\mathbf{t}} = \boldsymbol{\Phi} \mathbf{w}_{MP} + \mathbf{B}^{-1} (\mathbf{t} - \mathbf{y}) \quad (13)$$

These equations are basically equivalent to the solution of a generalized least-squares problem. After obtaining \mathbf{w}_{MP} , the hyperparameters α_i are updated with the i th posterior mean weight, λ_i and the i th diagonal element of the covariance Σ_{ii}

$$\alpha_i^{new} = \frac{\lambda_i}{w_i^2} \quad (14)$$

$$\lambda_i = 1 - \alpha_i \Sigma_{ii} \quad (15)$$

During the optimization process, many α_i will have large values, and thus, the corresponding model weights are pruned out, realizing sparsity. The optimization process typically continues until the maximum change in α_i values is below a certain threshold or the maximum number of iterations is reached. The corresponding \mathbf{x}_i of remained α_i is called relevance vector.

2.3 Space Weight

The spatial correlation of data points is an important part of information contained in HSI. The spatial information is also a remarkable basis for classification. However, such information is not carefully utilized in the above procedure. So, we decide to make a revise for the results gained from RVM. We assume the classification probability output of RVM is \mathbf{P} , the class number of HSI is NC , and the spatial weight matrix is \mathbf{W} . Then, the final result of HSI classification will be obtained as followed.

1) For \mathbf{P} , we first select the max probability of each pixel and set the class label as the corresponding one. Thus, we can get a preliminary classification result map \mathbf{Pmap} .

2) For pixel $\mathbf{Pmap}(i, j)$, $i, j = 1, \dots, NC$, we can count the numbers of pixels belonged to different classes in the eight adjacent areas. Hence the ratios of different classes around it are also gained, i.e., $\mathbf{W} = \text{diag}(W_1, W_1, \dots, W_{NC})$.

Then, modified probability \mathbf{P}_m could be achieved by

$$\mathbf{P}_m = \mathbf{W} \mathbf{P}_{i,j} \quad (16)$$

Where $\mathbf{P}_{i,j}$ is a probability vector of pixel (i, j) .

3) Execute the second step until every pixel of HSI is traversed. Therefore, we get the final modified probability output \mathbf{P}_m . Then, according to the operation of the first step, we can get the classification map with class labels.

3. EXPERIMENTS AND ANALYSIS

3.1 Data Description

A sample hyperspectral image which is taken over northwest Indiana's Indian Pine test site in June 1992 is used to test the proposed algorithm. The Indian Pine data consists of 145×145 pixels with 220 bands and 16 classes of substances. The number of spectral bands is initially reduced to 200 by removing bands, covering water absorption as well as noisy bands. The image of 10th band in hyperspectral image is shown in figure 1. The ground truth image is given in figure 2. All the pixels are labelled with numbers. One to sixteen stand for class indexes while zero means the pixel is not classified.



Figure 1. The image of 10th band

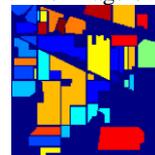


Figure 2. Ground truth image

3.2 Experiments and Results

We set the natural number labelled pixels as the sample set. The 35 percent samples are selected randomly as training samples to test the whole image. However, when comparing the performance of different algorithms, we still use the remaining 65 percent samples to calculate for comparative evaluation. RVM, SVM and proposed algorithm are compared on the same selected samples in this paper. Gaussian RBF kernel function is adopted as the basis function of RVM and SVM.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \quad (17)$$

Gaussian kernels parameter is 0.1. For SVM, the parameter cost is 125. Then, our classified result map is shown as figure 3. In addition, figure 4 shows the result map of RVM, and figure 5 shows for proposed algorithm. In the proposed method, $k = 20$ and the reduced dimension is 20 when ISOMAP is carried out.

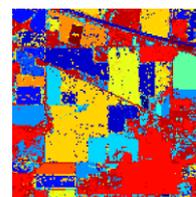


Figure 3. Classification results of SVM

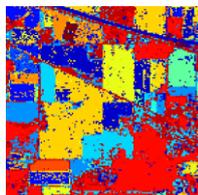


Figure 4. Classification results of RVM

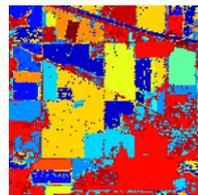


Figure 5. Classification results of proposed algorithm

3.3 Analysis

The Kappa coefficient and overall classification accuracy are often used to evaluate the classification results in the remote sensing image, we make a quantitative analysis on the experimental results. The results are shown in Table 1.

Algorithm	Kappa	OA
SVM	0.9023	91.44%
RVM	0.8681	88.46%
Proposed algorithm	0.9417	94.89%

Table 1. Kappa coefficient and overall classification accuracy

From the above indexes in table 1, we can find that the proposed method is superior to another two algorithms. Moreover, the highest accuracy of SVM and RVM is 97.95% (sixth class) and 94.63% (thirteen class) respectively, whereas that of method in this paper is 99.86% (thirteen class). In addition, 63.04% (first class), 41.30% (first class), and 83.67% (fourth class) are the lowest accuracies of SVM, RVM and proposed algorithm, respectively. Another point worth noting is that the support vectors of SVM is 3544, while the relevance vectors RVM and presented method are 541 and 514 respectively. Hence, the RVM classification is superior to the SVM classification in terms of sparsity.

Although spatial information is under consideration, we can still find that there are also some incorrectly classified points in block areas. This kind of result need to be further investigated and improved. Another problem is that the time consumption is increased with the method in this paper. Therefore, it is not suitable for real-time classification or the situation with time requirement.

4. CONCLUSION

Aiming at the problem that the classification accuracy of high dimensional data is not high, this paper presents a method that combine ISOMAP, RVM and spatial weight to reduce the dimensions of hyperspectral data and carry out classification. It is shown to provide higher classification accuracy compared with the SVM-based classification. With ISOMAP, the dimension is reduced and the information is well remained at the same time. With spatial weight, the result is improved. In general, this method can effectively enhance relevance vector machine model for the classification accuracy of hyperspectral data.

ACKNOWLEDGEMENTS

The authors would like to thank the websites (http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes, <http://www.disi.unitn.it/rslab>) for their opening data and programs. We would also like to thank the anonymous reviewers whose comments have significantly improved this paper.

REFERENCES

- Begüm Demir, Sarp Ertürk, 2007. Hyperspectral Image Classification Using Relevance Vector Machines. *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS*, 4(4), pp.586-590.
- Christopher M. Bishop, 2006. *Pattern Recognition and Machine Learning*. Springer, Singapore, pp. 345-356.
- C. I. Chang, 2003. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, USA: Academic, pp. 1-20.
- F. Melgani and L. Bruzzone, 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.*, 42(8), pp. 1778-1790.
- Freek van der Meer, 2005. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 8, pp. 3-17.
- M. E. Tipping, 2001. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine learning Research*, 1(3), pp. 211-244.
- Psorakis I, Damoulas T, Girolami M A, 2010. Multiclass relevance vector machines: sparsity and accuracy. *IEEE Trans Neural Netw*, 21(10), pp.1588-1598.
- Wnag Yiting, 2017. *Research on Hyperspectral Imagery Target Detection*. Xi'an, pp.1-18.
- Tenenbaum J B, de Silva V, Langford J C, 2000. A global geometric frame work for nonlinear dimensionality reduction. *Science*, 290(5500), pp. 2319-2323.