CLASSIFICATION OF REMOTELY SENSED IMAGE USING RELEVANCE VECTOR MACHINE

A.Kalarani, G.viji, S.Ramprakash

1Assistant Professor, P.S.R.Rengasamy college of engg for women, Sivakasi.
2Assistant Professor, P.S.R.Rengasamy college of engg for women, Sivakasi.
3Lecturer, M.Kumarasamy College of Engg, Karur.

Abstract— This paper introduces a remotely sensed image classification method based on relevance vector machines (RVMs). The features of the remotely sensed image are extracted and the classification is done[4] with the help of those features. It is shown that approximately the good classification accuracy is obtained using RVM-based classification, with a significantly smaller relevance vector rate and, therefore, much faster testing time. This feature makes the RVM-based classification approach more suitable for applications that require low complexity and, possibly, real-time classification.

Index Terms — Classification, remotely sensed image, Bayesian learning, relevance vector machines (RVMs).

I. INTRODUCTION

In the recent years, relevance vector machines (RVMs) have been successfully used in many application domains. In particular, the RVM constitutes a Bayesian approximation for solving generalized linear classification and regression models[1]. This method not only provides accurate predictions but also force sparsity (simplicity) of the method, and can produce confidence intervals for the predictions. Good trade-offs between accuracy and sparseness of the solution has been observed in many application domains. In the field of remote sensing, the use of RVM has been recently introduced for the prediction of biophysical parameters. Being a kernel-based method, the key point for obtaining good RVM classifiers is the definition of a suitable kernel function that can properly represent relations (similarities) among samples (pixels).

The advantages of the RVM are probabilistic predictions, automatic estimations of parameters, and the possibility of choosing arbitrary kernel functions. Most importantly, RVM classification results[9] in fewer relevance vectors (RVs), classification can be carried out much faster with the RVM. For example, the RVM has been used for the detection of micro calcification clusters in digital mammograms, and it has been shown that the RVM classifier is much more suitable for real-time processing and reduces the computational complexity while maintaining similar detection accuracy. It is proposed in this letter to utilize the RVM for classification of remotely sensed images. This feature makes the RVM based classification approach more suitable for applications that require low complexity and possibly, real-time classification.

II. PROPOSED METHODOLOGY

The proposed methodology classifies the remote sensed image based on RVM algorithm. In the first stage the remote sensed image is transformed using DWT. The approximated image is then chosen. The features of the approximated image were extracted .The extracted features were classified into
i)statistical features
ii)textural features
The statistical features include i) mean ii) variance and iii) standard deviation. The textural features include i) energy ii) entropy iii) contrast and iv) homogeneity. The extracted features were taken as training and testing samples. The training and testing samples were classified using RVM algorithm and the performance were measured[12].

III. RVM CLASSIFICATION

Supervised learning techniques make use of a training set that consists of a set of sample input vectors \( \{ x_n \}_{n=1}^{N} \) together with the corresponding targets \( \{ t_n \}_{n=1}^{N} \). The targets are basically real values in regression tasks or class labels in classification problems. It is typically desired to learn a model of the dependency of the targets on the inputs from the training set, so that accurate predictions of \( t \) can be made.
for previously unseen values of $x[8]$. Commonly, these predications can be based on some function $y(x)$ defined over the input space in the form of

$$ y(x; w) = \sum_{i=1}^{M} w_i \Phi(x) = w^T \Phi(x) \quad (1) $$

as a linearly weighted sum of $M$ (generally nonlinear and fixed) basis functions $\Phi(x) = (\Phi_1(x), \Phi_2(x), \ldots, \Phi_M(x))^T$.

Although this model is linear in the parameters (or weights), $w = (w_1, w_2, \ldots, w_M)^T$ it can still be highly flexible as the size of the basis set $M$ can be effectively large. Learning is basically the process of inferring the function or, equivalently, the parameters of the function $y(x)$. In this context, it is desired to estimate reasonable values for the parameters (or weights), $w = (w_1, w_2, \ldots, w_M)^T$. Given a set of $N$ corresponding training pairs $\{x_n, t_n\}_{n=1}^{N}$, the objective is to find values for the weights $w = (w_1, w_2, \ldots, w_M)^T$, such that $y(x)$ generalizes well enough to new data, yet only a few elements of $w$ are nonzero[5]. Having only a few nonzero weights facilitates a sparse representation with the advantage of providing fast implementation.

The RVM introduces a prior over the model weights governed by a set of hyper parameters $\alpha$, in a probabilistic framework. One hyper parameter is associated with each weight, and the most probable values are iteratively estimated from the training data[1]. The most compelling feature of the RVM is that it typically utilizes significantly fewer kernel functions, while providing a good performance. For two-class classification, any target can be classified into two classes such that $t_n \in \{0,1\}$. A Bernoulli distribution can be adopted for $p(t|x)$ in the probabilistic framework because only two values (0 and 1) are possible. The logistic sigmoid link function $\sigma(y) = 1/(1 + e^{-y})$ is applied to $y(x)$ to link random and systematic components, and generalize the linear model.

Following the definition of the Bernoulli distribution, the likelihood is written as

$$ p(t/w) = \prod_{n=1}^{N} \sigma(y(x_n;w))^t_n [1 - \sigma(y(x_n;w))]^{1-t_n} \quad (2) $$

for the targets $t_n \in \{0,1\}$. The likelihood is complemented by a prior over the parameters(weights) in the form of

$$ p(w/\alpha) = \prod_{i=1}^{N} \frac{\alpha_i}{2\pi} \exp\left(-\frac{\alpha_i w_i^2}{2}\right) \quad (3) $$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)^T$ shows the hyper parameters introduced to control the strength of the prior over its associated weight[3]. Hence, the prior is Gaussian, but conditioned on $\alpha$. For a certain $\alpha$ value, the posterior weight distribution conditioned on the data can be obtained using Bayes’ rule, i.e.,

$$ p(w/t,\alpha) = \frac{p(t/w)p(w/\alpha)}{p(t/\alpha)} \quad (4) $$

where $p(t/w)$ is the likelihood, $p(w/\alpha)$ is the prior, and $p(t/\alpha)$ is referred to as evidence. The weights cannot be analytically obtained, and therefore, a Laplacian approximation procedure is used.1) Since $p(t/w,\alpha)$ is linearly proportional to $p(t/w) \times p(w/\alpha)$, it is possible to aim to find the maximum of

$$ \log\left\{ p(t/w)p(w/\alpha) \right\} = \sum_{n=1}^{N} \left[ t_n \log y_n + (1 - t_n) \log(1 - y_n) \right] - \frac{1}{2} w^T A w \quad (5) $$

for the most probable weights $W_{MP}$, with $y_n = \sigma(y(x_n;w))$ and $A = \text{diag}(\alpha_0, \alpha_1, \ldots, \alpha_N)$ being composed of the current values of $\alpha$. This is a penalized logistic log-likelihood function and requires iterative maximization. The iteratively reweighed least-squares algorithm can be used to find $W_{MP}[6]$. The logistic log-likelihood function can be differentiated twice to obtain the Hessian in the form of

$$ \nabla w \nabla^T \log\left\{ p(t/w,\alpha) \right\} \mid_{w_{MP}} = - (\varphi^T B \varphi + A) \quad (6) $$

where $B = \text{diag}(\beta_1, \beta_2, \ldots, \beta_N)$ is a diagonal matrix with $\beta_n = \sigma(y(x_n;w))/(1 - \sigma(y(x_n;w)))$, and $\Phi$ is the ‘design’ matrix with $\Phi_{nm} = K(x_n, x_m)$ and $\Phi_{nn} = 1$. This result is then negated and inverted to give the covariance $\Sigma$, as shown as follows[12], for a Gaussian approximation to the posterior over weights centered at $W_{MP}$.

$$ \Sigma = (\Phi^T B \Phi + A)^{-1}. \quad (7) $$

In this way, the classification problem is locally linearized around $W_{MP}$ in an effective way with

$$ W_{MP} = \Sigma \Phi^T B^{-1} \quad (8) $$

$$ t = \Phi W_{MP} + B^{-1}(1 - y). \quad (9) $$

These equations are basically equivalent to the solution of a generalized least-squares problem. After obtaining $W_{MP}$, the
Hyper parameters $\alpha_i$ are updated using $\alpha$, 
$\alpha_i^{new} = \lambda_i / w_i^2$ where $w_i$ is the $i$th posterior mean weight, and $\lambda_i$ is defined as $\lambda_i = 1 - \alpha_i \Sigma_{ii}$, where $\Sigma_{ii}$ is the $i$th diagonal element of the covariance, and can be regarded as a measure of how well determined each parameter $w_i$ is by the data[15]. During the optimization process, many $\alpha_i$ 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 $\alpha_i$ values is below a certain threshold or the maximum number of iterations is reached.

### III. Experimental Results

In this section, the proposed RVM classifier is tested on an urban image of the area of Pavia, Italy.

![Fig (a)](image1) ![Fig (b)](image2)

Fig. (a) RGB composition of Pavia image, and b) groundtruth. This image was acquired by the DAIS 7915 airborne imaging spectrometer of DLR. This is a challenging urban classification problem dominated by directional features and relatively high spatial resolution. Different values of the width for the kernel were tried exponentially.

The most popular kernels used in RVM are the linear, polynomial, and radial basis function (RBF) kernels. The RBF kernel typically shows a performance and is therefore employed in the provided results. Note that $\gamma$ serves as an inner product coefficient for the polynomial kernel, whereas it determines the RBF width in the case of the RBF kernel.

**Linear kernel**

$$K(x_i, x_j) = x_i \cdot x_j$$

**Polynomial kernel**

$$K(x_i, x_j) = (\gamma x_i \cdot x_j)^d$$

**RBF kernel**

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

The accuracy and the relevance vector for the extracted features (homogeneity and contrast) are tabulated as

<table>
<thead>
<tr>
<th>Model</th>
<th>Features</th>
<th>AC</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVM</td>
<td>homogeneity</td>
<td>96</td>
<td>5</td>
</tr>
<tr>
<td>RVM</td>
<td>contrast</td>
<td>97</td>
<td>11</td>
</tr>
</tbody>
</table>

The RV plots for the two class problem $\{0,1\}$ for the features homogeneity and contrast are shown in Figures 1 and 2 respectively.

![Fig. 2](image3)

**Fig. 2**. Classification maps obtained for a two-class problem $\{0,1\}$ for the feature homogeneity. Red and blue dots indicate the classes $\{0,1\}$, red dots point out the relevant vectors (RVs), the red line represents the classification boundary, and the grey lines are the confidence intervals at $p = 0.25$ and $p = 0.75$.  

All Rights Reserved © 2012 IJARCSEE
Fig. 3. Classification maps obtained for a two-class problem for the feature contrast. Red and blue dots indicate the classes /0,1/, red dots point out the relevant vectors (RVs), the red line represents the classification boundary, and the grey lines are the confidence intervals at $p = 0.25$ and $p = 0.75$.

IV. CONCLUSION

RVM-based image classification provide good classification accuracy, with a significantly smaller RV rate and therefore, much faster testing time. The most evident and compelling results are its accuracy and sparseness. RVM-based classification approach is more suitable for applications that require low complexity and, possibly, real-time classification.

REFERENCES


Kalari Athilingam completed her B.Engg. degree in Electronics and Communication Engineering from Anna University, Chennai, in 2008 and the Master of Engg. degree from Anna University, Tirunelveli, in 2010. From June 2010 to till now, She is working in P.S.R. Rengasamy College of Engg for women, Sivakasi. Her research area includes Digital Electronics, Digital Image processing, Antenna. Communication theory. She has been attended several workshops and conferences in various engg colleges.

Viji Gurusamy received the B.Engg. degree in Electronics and Communication Engineering from Anna University, Chennai, in 2008 and the Master of Engg. degree from Anna University, Tirunelveli, in 2010. From June 2010 to May 2012, She was worked in M. Kumarasamy College of Engg, Karur. Now she is currently working in P.S.R. Rengasamy College of Engg for women, Sivakasi. She had attended four international conferences and one national conference in various colleges. Her research area includes Digital Signal processing, Digital Image processing, Digital Communication.

Ramprakash Subburam received the B.Engg. degree in Electronics and Instrumentation Engineering from Anna University, Chennai, in 2009 and doing Master of Engg. degree in Anna University, coimbatore. He has been worked as an Instrumentation Site Engineer in Micotec Engineers and contractors (A sub contractor to Yokogawa India ltd) to Empee Cogen power plant, Edaiakal, Tirunelveli District from may 2009. From June 2010 to till now, he is working in M. Kumarasamy College of Engg, Karur. His research area includes wireless communication, Bio-medical instrumentation, process control, Digital Image processing.