High Emphatic Feature Selection Classification Algorithm for SVM in Remote Sensing Images

T. Balaji [1], M. Sumathi [2]

[1] Assistant professor, Dept. of Computer Science, Govt. Arts College, Melur - 625106, India
bkmd_gacm@rediffmail.com
[2] Associate Professor, Dept. of Computer Science Sri Meenakshi Govt. College for Women, Madurai - 625 002, India
sumathivasagam@gmail.com

Abstract - In image classification, an image is classified according to its visual content. An important application is image retrieval-searching through an image dataset to obtain those images with particular visual content. The SVM is a powerful technique for remote sensing image classification. In this paper, we propose a simple procedure or a function which is usually gives a reasonable results. The main goal of this paper includes: **1**. Train a visual classifier for different image classes **2.** Achieve the performance of the classifier **3.** Visual content used for Mapping (feature in image) **4**. Obtain the train image data. The SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. Note that this SVM approach guide to achieve the highest accuracy. Also, we intent to solve challenging or difficult problems in remote sensing image data sets.

Keywords: SVM, RBF, Hyperplane, Polinomial Kernel, Multilayer Perception.

1. Introduction

The foundations of Support Vector Machines (SVM) have been developed in 1995 and are gaining popularity due to many attractive features, and promising empirical performance. The SVM is a relatively recent addition to the remote sensing image classification. This is a new generation learning system based on the latest advances in statistical learning theory. The SVM is useful in two perspectives [1]: First, Support vector learning is based on some beautiful simple ideas and provides a clear intuition of learning from example. Second, it can lead to high performance in practical applications in image classification, hand written character recognition, image classification, and remote sensing imagery classification fishery classification and so on.
Although SVM is considered easier to use than Neural Networks. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one “target value" (i.e. the class labels) and “several attributes" (i.e. the features or observed variables). The SVM also belongs to the class of supervised learning algorithms in which learning machine is given a set of examples with the associated labels. The SVM is to construct a hyperplane that separates two classes or extended to multi-class problems [1]. The SVM algorithm tries to achieve maximum separation between the classes. The SVM for separating the classes with a large margin minimize a bound on the expected generalization error. This means that new examples (data points with unknown class values) arrive for classification, the chance of making error in our prediction based on the learned classifier should be minimum.

2. Proposed SVM Classifier Features

The SVM uses the concept of “maximum margin”. The two planes parallel to the classifier and which passes through one or more points in the image data set is called **bounding planes [1]**. The distance between these bounds places is called margin and by SVM learning, which means finding a hyperplane with maximize this margin. The structure containing information about the trained SVM classifier, including the following fields. A maximum margin classifier is shown in the following figure 1.

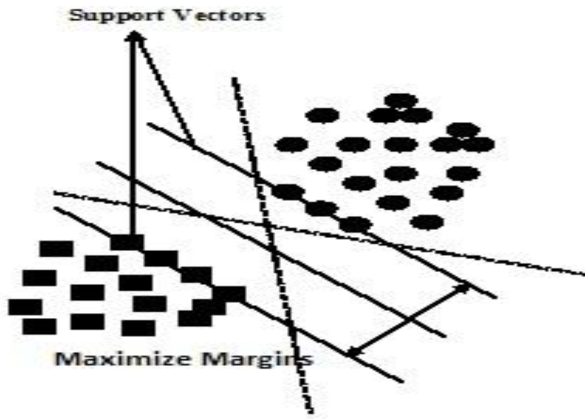

Fig. 1. A Maximum Margins Classifier

The points falling on the bounding planes are called **support vectors**. These points to play a crucial role in the theory and hence the name support vector machines. By a'Machine', means as the algorithm. The SVM method is either treated as linear classifier or non-linear classifier. The data points are clustered such that the two classes are linearly separable. Hence the classifier is called linear classifier. It may have to tolerate large training error. The points on the images are clustered such that the two hyperplane classes are not linearly separable. Hence the classifier is called non-linear classifier. It may be used for two dimensional and three dimensional input points.

The key factor of SVM is penalty parameter C and Kernel function parameters. The most widely used kernels functions are polynomial kernels and RBF kernels (Radial Bias Function). Originally SVM was developed to perform binary classification. However, most of the classification problems involved in remote sensing is more than two classes. Two simple strategies follow, namely one-against-one and one-against-rest respectively. The one-against-one is to construct a classifier for each pair of classes, machine is trained to classify a pixel labeled having the most visible feature in a class. Another one-against-rest is to break the multi-class case into n number of two-classes machine is trained to classify one class against all other classes.

3. Understanding SVM

The SVMs were developed to solve the classification problem, but recently they have been extended to the domain of regression problems.

3.1. SVM With Separable Data

The SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class. The best hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points. The support vectors are the data points that are closest to the separating hyperplane; these points are on the boundary of the slab. The following figure illustrates these definitions, with **"+"** indicating data points of type 1, and **" –"** indicating data points of type –1.

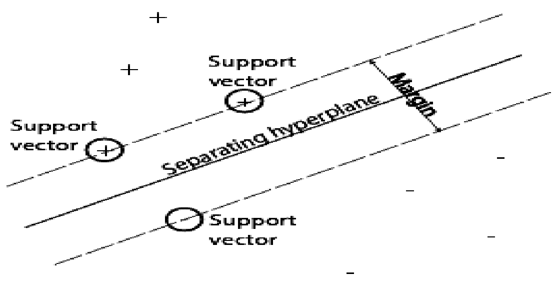

Fig. 2. SVM with Separate Hyperplane

The data for training is a set of points (vectors) x_i along with their categories y_i. For some dimension d, the $x_i \in R^d$, and the $y_i = \pm 1$. The equation of a hyperplane is $\langle w,x\rangle + b = 0$, where $w \in R^d$, $\langle w,x\rangle$ is the inner (dot) product of w and x, and b is real. The following problem defines the *best* separating hyperplane. Find w and b that minimize $||w||$ such that for all data points (x_i,y_i),

$$y_i(\langle w,x_i\rangle + b) \geq 1.$$

The support vectors are the x_i on the boundary, those for which $y_i(\langle w,x_i\rangle + b) = 1$. It is computationally simpler to solve the dual quadratic programming problem. To obtain the dual prominent factor:

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle, \qquad (1)$$

In which we maximize over $\alpha_i \geq 0$. In general, many α_i are 0 at the maximum. The nonzero α_i in the solution to the dual problem define the hyperplane, which gives w as the sum of $\alpha_i y_i x_i$. The data points x_i corresponding to nonzero α_i are the *support vectors*. The derivative of L_D with respect to a nonzero α_i is 0 at an optimum. This gives $y_i(\langle w,x_i\rangle + b) - 1 = 0$.

3.2. Nonseparable Data

This method data might not allow for a separating hyperplane. In that case, SVM can use a soft margin, meaning a hyperplane that separates many, but not all data points.

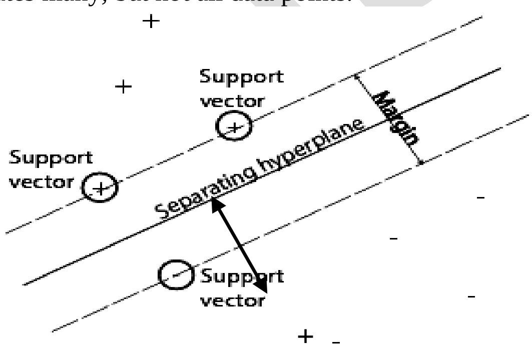

Fig. 3. SVM with Nonseparable Hyperplane

There are two standard formulations of soft margins. Both involve adding slack variables si and a penalty parameter C. The first and second norm of nonseparable data formulation factor is:

$$\min_{w,b,s}\left(\frac{1}{2}\langle w,w\rangle + C\sum_i s_i\right) \quad \text{and} \quad \min_{w,b,s}\left(\frac{1}{2}\langle w,w\rangle + C\sum_i s_i^2\right) \qquad (2)$$

3.3. Nonlinear Transformation with Kernels

Some binary classification problems do not have a simple hyperplane as a useful separating criterion. For those problems, there is a variant of the mathematical approach that retains nearly all the simplicity of an SVM separating hyperplane. This approach uses these results from the theory of reproducing kernels: There is a class of functions $K_n(x,y)$ with the following property. There is a linear space S and a function φ mapping x to S such that

$K_n(x,y) = \langle\varphi(x),\varphi(y)\rangle$.

The major kernel functions are Gaussian Radial Basis Function (RBF), linear, polynomial and sigmoid functions:

Polynomials: For some positive integer d,

$$Kn(x,y) = (1 + \langle x,y\rangle)^d$$

Radial Basis function: For some positive number σ,

$$K_n(x,y) = \exp(-\langle(x-y),(x-y)\rangle/(2\sigma^2))$$

Multilayer Perceptron (neural network): For a positive number p_1 and a negative number p_2,

$$K_n(x,y) = \tanh(p_1\langle x,y\rangle + p_2)$$

The final **decision function** is defined as

$$f(x) = \text{sign}\left(\sum y_i\alpha_i K(x_i,x) + b\right) \quad (3)$$

4. Steps for Predictive Classification Accuracy of SVM Classifier

The accuracy of classification method is the ability of the method to correctly determine the classification of randomly selected data instance. It may be expressed as the probability of correctly classifying unseen data. The optimistic predictions are often made regarding the accuracy of the classification method. We propose that beginners try the following procedure to improve the classification accuracy of SVM classifier:

4.1. To Train the Visual Classifier

The frequency of each visual word of classification is data preparation. The data provided in the folder, data consist of image and pre-computed feature vectors for each image. The data consists of different image classes. The svmtrain function uses an optimization method to identify support vectors s_i, weights α_i, and bias b that are used to classify vectors x according to the following equation:

$$c = \sum_i \alpha_i k(s_i, x) + b, \quad (4)$$

where k is a kernel function. In the case of a linear kernel, k is the dot product. If $c \geq 0$, then x is classified as a member of the first group, otherwise it is classified as a member of the second group. The resulting SVMstruct contains the optimized parameters from the SVM algorithm enabling to classify new data. The classifying new data with an SVM classifier using **svmclassify** function. The resulting vector of new classes to represents the classification of each row in new data.

4.2. Performance of SVM Classifier

The performance of the SVM classifier score to rank all the remote sensing test images. Note that the bias term is not needed for this ranking, only classification feature vector should be observed. Now we will measure the retrieval performance quantitatively by computing precision methods. The precision method is computed by varying the threshold on the classifier. In order to assess the retrieval performance by a single number, the average precision (AP) is often computed.

4.3. Visual Content used for Feature Mapping

The **svmtrain** function uses an optimization method to identify support vectors $\mathbf{S_i}$, weights $\mathbf{W_i}$ and bias b that are used to classify vector x according to the following equation:

$$\mathbf{C = \sum_{i=0}^{n} W_i \, . \, K\,(S_i\,,\,x) + b} \qquad (5)$$

Where K is a kernel function. If $C \geq 0$, then x is classified as a member of the first group, otherwise it is classified as a member of the second group. The process of nonlinear mapping illustration into feature space is shown in the following figure 2. We may feel that the process involves heavy computation especially when we map the data which itself may be of high dimensional feature space. Mapping of all computations will be done in the space itself.

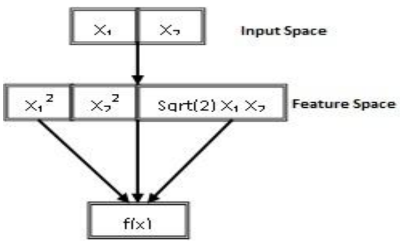

Fig. 4. Mapping of Input Feature Space

4.4. Obtain the Trained Image Data

This formulation leads to an extremely fast and simple algorithm for generating linear or nonlinear classifier that is obtained by solving the remote sensing image and optimize its retrieval performance. We want to obtain train image data, all images are automatically normalized to a standard size and descriptors are saved for each new image added in the data folder or directory.

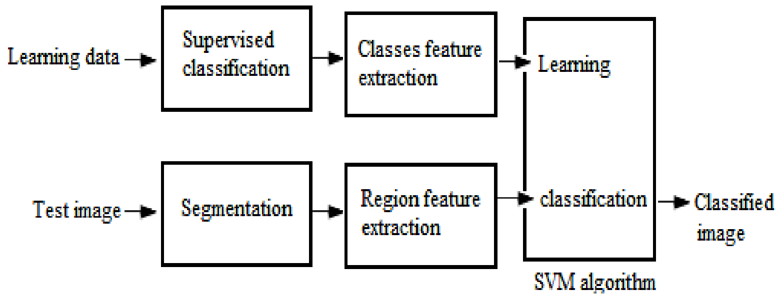

Fig. 5. Feature Selection and SVM Classification

5. Advantages of Support Vector Machine (SVM) Methods

•**SVM is a modern outgrowth of artificial neural networks.** Support Vector Machine methods are close cousins to neural networks. In fact, a SVM method using a sigmoid kernel function is equivalent to a two-layer, feed-forward neural network.

•**Highly accurate models.** The SVM approach guide to achieve the highest **accuracy** and also intent to solve challenging or difficult problems in remote sensing image data sets.

•**Classification and Regression analyses.** The implementation of SVM method supports binary and multi-class classification problems as well as regression. It also implements the most popular kernel functions including radial basis functions, sigmoid, polynomial and linear.

•**Automatic grid search and pattern search for optimal parameters.** The accuracy of SVM method depends on selecting appropriate parameter values. It provides an automatic grid and pattern search facility that allows it to iterate through ranges of parameters and perform cross-validation to find the optimal parameter values.

•**Model building performance.** The implementation of SVM is capable of handling very large problems. Kernel matrix row caching, shrinking heuristics to eliminate outlying vectors and this type algorithm are used to boost the speed of modeling.

•**Missing value substitution.** If there are scattered missing values for predictor variables, it can replace those missing values with median values so that the case can be salvaged non-missing variable values used to the maximum extent.

•**V-fold cross validation.** It provides V-fold cross validation both during the search process to select the optimal **parameters and as a verification method for the final model.**

6. Experimental and Implementation Issues

Many kernel mapping functions can be used – as a probably an infinite number. But a few kernel functions have been found to work well in for a wide variety of applications. The default and recommended kernel function is the Radial Basis Function (RBF).

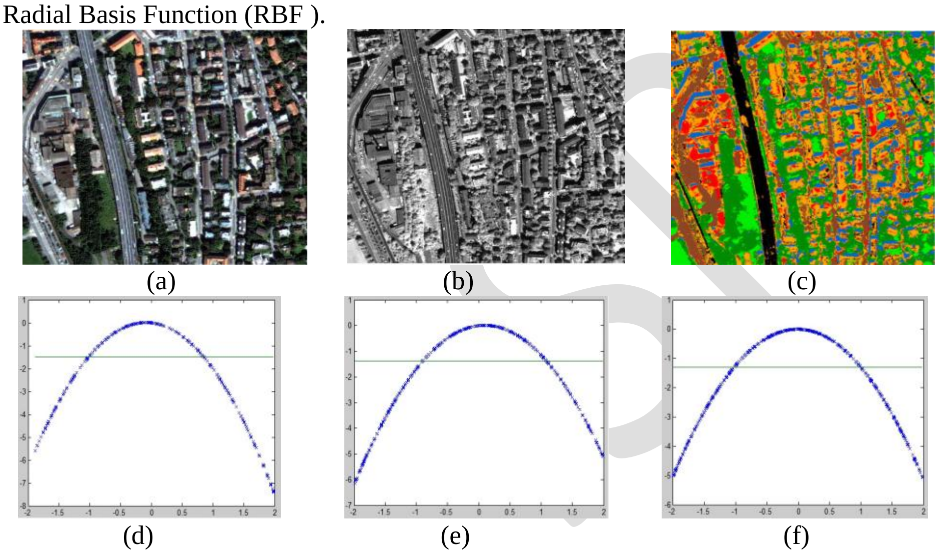

Fig. 6. a). Original commercial Earth Observation Satellite image, b). RGB to Gray Image, c). SVM Classified Image, d). Noise Factor 0.8, e). Noise Factor 0.9, f). Noise Factor 1.0 with Randomized Pixels

7. Conclusion

There are several SVM algorithms are constantly appearing year by year. Now-d-days, the SVM algorithm tries to achieve maximum separation between the classes. Comparing with the other classification approaches, SVM can achieve high classification accuracy with small training samples and is less dependent on the statistical distribution of the data. The SVM method to classifiers data by finding the best hyperplane that separates all data points of one class from those of the other class. The plenty of researchers to prove the SVM perform well in remote sensing image classification accuracy in order to boost the remote sensing applications. The different classifier has different properties that make it better suited to particular types of classifications effectively into an ensembles is worth for further studying. The SVM approach always resulted in better classification accuracy with more significant improvement compared to the non-feature-selection approach. The highest overall classification accuracy of 96.97% was obtained using the SVM method for classifying the combined dataset of original remote sensing image.

Discussions

For the evaluation of the developed SVM approach, remote sensing images were used. Comparison with manual segmentation and got good results, we much improved efficiency of land resource investigation work when used it in interactive resolution image segmentation. Overall, the SVM classification approach was found very promising for Image Analysis. It has been shown that it can produce comparable or even better results than the Nearest Neighbor for supervised classification. A very good feature of SVMs is that only a small training set is needed to provide very good results, because only the support vectors are of importance during training. The Support vector approach, with significantly higher classification accuracy and much fewer input features. The highest classification accuracy achieved was 96.47% was obtained using the SVM method for classifying the combined dataset of remote sensing polarized image.

Reference

[1] T.Balaji and M.Sumathi, "Relational Features of Remote Sensing Image Classification using Effective K-Means Clustering", IJoART, pp. 103 – 107, August 2013.

[2] C.C. Chang and C-J Lin., "A Library for SVM", May 2013.

[3] M. Pal, and P. M. Mather, "Support Vector Machines for classification in remote sensing", International Journal of Remote Sensing, vol. 26, pp. 1007–1011, 2012.

[4] Chih-Wei Hsu, "A Practical Guide to Support Vector Classification", pp. 1-16, April 2012.

[5] C. Cortes and V. Vapnik, "Support Vector Networks", Machine Learning, Vol-20, pp. 273-297, 2012.

[6] Yan LI, Li YAN, "Remote Sensing image Classification Development in the Past Decade", Wuhan university, pp. 1-6, 2011.

[7] Chih-Wei Hsu, "A Practical Guide to Support Vector Classification", pp. 1-16, April 2011.

[8] Steve R. Gunn, "Support Vector Machines Classification", University of Southampton, May 2008.

[9] S.S. Keerthi and C-J Lin., "Asymptotic Behaviors of SVM with Gaussian Kernel", Neural Computation, Vol-15, pp. 1667-1689, 2011.

[10] Andrea Vedaldi and Andrew Zisserman, "Image Classification Practical", pp. 1-7, 2011.

[11] Lu, D. and Weng, Q., "A survey of Image Classification Methods and Techniques for Improving Classification Performance", International journal of remote Sensing, pp. 823-870, 2010.