Satellite Image Segmentation using RVM and Fuzzy clustering

Harsh Pareek

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela – 769 008, India

Satellite Image Segmentation using RVM and Fuzzy clustering

Dissertation submitted in

May 2014

to the department of

Computer Science and Engineering

of

National Institute of Technology Rourkela

in partial fulfillment of the requirements

for the degree of

Master of Technology

by

Harsh Pareek

(Roll 212CS1096)

under the supervision of

Ratnakar Dash

Department of Computer Science and Engineering

National Institute of Technology Rourkela

Rourkela – 769 008, India

dedicated to my Parents and Brother...

DECLARATION

I, Harsh Pareek (Roll No. 212CS1096) understand that plagiarism is defined as any one or the combination of the following

1. Uncredited verbatim copying of individual sentences, paragraphs or illustrations (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Uncredited improper paraphrasing of pages or paragraphs (changing a few words or phrases, or rearranging the original sentence order).

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did or wrote what. (Source: IEEE, the Institute, Dec. 2004)

I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the thesis may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Date

Harsh Pareek
212CS1096

Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela-769 008, India. www.nitrkl.ac.in

Dr. Ratnakar Dash
Asst. Professor

May , 2014

Certificate

This is to certify that the work in the thesis entitled *Satellite Image Segmentation using RVM and Fuzzy clustering* by *Harsh Pareek*, bearing roll number 212CS1096, is a record of research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of *Master of Technology* in *Computer Science and Engineering*. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Ratnakar Dash

Acknowledgement

This dissertation, though an individual work, has benefited in various ways from several people. Whilst it would be simple to name them all, it would not be easy to thank them enough.

The enthusiastic guidance and support of *Asst. Prof. Ratnakar Dash* inspired me to stretch beyond my limits. His profound insight has guided my thinking to improve the final product. My solemnest gratefulness to him.

I am also grateful to *Prof. Banshidhar Majhi* for his ceaseless support throughout my research work.
I extend my thanks to our HOD, *Prof. S. K. Rath* for his valuable advices and encouragement.

I am really thankful to my all friends.

Last, but not the least, I would like to dedicate this thesis to my family, for their love, patience, and understanding.

Harsh Pareek

Abstract

Image segmentation is common but still very challenging problem in the area of image processing but it has its application in many industries and medical field for example target tracking, object recognition and medical image processing. The task of image segmentation is to divide image into number of meaningful pieces on the basis of features of image such as color, texture.

In this thesis some recently developed fuzzy clustering algorithms as well as supervised learning classifier Relevance Vector Machine has been used to get improved solution. First of all various fuzzy clustering algorithms such as FCM, DeFCM are used to produce different clustering solutions and then we improve each solution by again classifying remaining pixels of satellite image using Relevance Vector Machine (RVM) classifier.

Result of different supervised learning classifier such as Support Vector Machine (SVM), Relevance Vector Machine (RVM), K-nearest neighbors (KNN) has been compared on basis of error rate and time.

One of the major drawback of any clustering algorithm is their input argument that is number of clusters in unlabelled data. In this thesis an attempt has been made to evaluate optimal number of clusters present in satellite image using DAVIES-BOULDIN Index.

***Keywords*:** Satellite Image Segmentation, Fuzzy Clustering algorithms, Fuzzy C-mean clustering, Differential Evolution Fuzzy Clustering algorithm, SVM, RVM, KNN, Evalution of optimal number of clusters, Davies-Bouldin Index

Contents

List of Figures

List of Tables

Chapter 1

Inroduction

Images plays very important role in human life. An image can be viewed as a two dimensional signal, it can be defined by using a mathematical function f(x,y) where x and y are horizontal and vertical coordinates respectively. Value of f(x,y) function is nothing but pixel value at point x and y. Digital image processing is a process of manipulating, analyzing and processing digital images. Input and output of digital processing system are images it process image and gives output. First of all a image is captured using device such as camera, satellite, microscope then it is pass as input to digital processing system then after processing it is given as output. [1]

Images can also be classified into different types depending upon their brightness and color intensities. The different types of images are as follows:

1. **Binary Image**

   Binary images are also known as black and white images. These images takes only two values either 0 or 1 where 0 represent black and 1 represent white.

2. **Grayscale Image**

   In grayscale image each pixel represent amount of light or brightness information. Each pixel have size of a byte and ranges from 0 to 255 where 0 represent black and 255 represent white.

3. **Colour Image**

   In color image each pixel has three values one for each RGB (Red Green Blue), which is generally used colour space, other colour spaces are HSV, CMYK. Thus in colour image each pixel is of size 3 bytes.

Digital image processing have its applications in various fields but requirement of large amount of memory and high processing speed can be considered as its disadvantage. Some of its advantages are: [1]

1. **Intelligent Transportation System**

   For automating the process of recognition of number plate and traffic signal digital image processing is used. Thus digital image processing system helps in creating intelligent transportation system and this process reduces crimes as well as violation of traffic rules.

2. **Remote Sensing**

   Processing of image taken from remote areas helps in flood control,city planing,agricultural production monitoring.

3. **Military and Defence**

   Digital Image Processing has been widely used in the field of military and defence. Its applications in defence are such as target detection and tracking, vehicle navigation, missile guidance.

4. **Medical**

   X-Ray, Ultrasound are widely used in the field of medical from the past many years. Application of DIP in field of medical such as heart disease, lung disease identification.

5. **Tracking Object in Motion**

   This application helps in tracking moving objects for example moving vehicle. Two methods to track an object are:

Motion based tracking and

Recognition based tracking

6. **Automatic Visual Inspection System**

   This application of Digital Image Processing is used in industries for improvement of product quality and checking wether product is faulty or not.

Various methodologies applied on image [1].

(i) Image Enhancement

    Image enhancement is technique to highlight certain areas of interest in an image. Among other processes it is simplest one.

(ii) Image Restoration

    Image Restoration also deals with improving the appearance of image. But unlike image enhancement it uses mathematical and probabilistic model.

(iii) Image Compression

    Compression techniques are used for reducing storage requirement of image.

(iv) Image Segmentation

    Segmentation deals with partioning image into objects so that each object can be identified individually.

The rest of the chapter is organized as follows. The introduction of Satellite Image Segmentation and its applications are given in Section 1.1. Some of the existing litratures are reviewed in Section 1.2. The research motivation and its objective are formally stated in Section 1.3. Finally Section 1.4 outline the layout of thesis.

1.1 Satellite Image Segmentation

Image segmentation is common but still a challenging problem in the area of image processing but it has its applications in many areas such as object tracking, object recognition.Application of satellite image segmentation is in the areas such as climate studies, examining marine enviroment, city planing, agricultural production monitoring. The main task in satellite image segmentation is classification of pixels into homogeneous regions based on some features of image.

Image segmentation problem can be viewed as unsupervised classification technique. For satellite image segmentation fuzzy clustering technique is better than hard clustering technique as pixels in satellite image contain large amount of imprecision and uncertainty. Therefore fuzzy clustering is appropriate for satellite image segmentation. The task of clustering is to divide pixels into groups such that pixels in the same cluster are more similar than pixels belonging to other clusters.

Main disadvantage of these clustering algorithms is to pass number of clusters as argument. In this thesis Davies-Bouldin index is used to evaluate optimal number of clusters present in an image. After that various fuzzy clustering algorithms like fuzzy c-mean, Differential evolution fuzzy c mean clustering (DeFCM) and modified differential evolution fuzzy c mean (MoDEFCM) are applied on image using evaluated number of clusters. Then using the result of clustering algorithms various supervised classifiers such as relevance vector Machine (RVM), Support Vector Machine (SVM), K-nearest neighbor(KNN) are used and their results are compared.

The fuzzy c-means (FCM) algorithm produce partition matrix U(x) which is probability of pixels belonging to cluster with constraint of minimizing measure j_m. However, recently proposed modified differential evolution based fuzzy clustering (MoDEFC) have used the Xie-Beni index while in support vector machine ensemble fuzzy clustering (SVMeFC) algorithm SVM classifier is applied on result of previously mentioned fuzzy clustering algorithms to produce good result. [2–4]

1.1.1 Image Segmentation Technique

Image Segmentation is a technique in which image is divided into subparts or regions. Each pixel of an image belongs to one of the class. Pixels belonging to same class or group are more similar than pixels belonging to different class. Image segmentation is done on the basis of some similarity measure. Image segmentation is natural but it is critical and challenging step in the field of image processing. [1] There are two approaches for Image Segmentation:

(i) Discontinuity base image segmentation.

(ii) Similarity base image segmentation.

(i) Discontinuity base image segmentation.

In Discontinuity Based Image Segmentation we use mask approach for detection of isolated points, lines and edges. Formulae used: R = W f(x, y) where w is weight of mask and f(x,y) is intensity at position x,y.

(a) Detection of isolated points For the detection of isolated point mask is applied on image and value of R is calculated. Then the calculated R value is compared with T, T is non negative threshhold, if value of R is greater than threshhold then we say an isolated point is detected at location x,y.

(b) Detection of lines

For the detection lines there are four masks used one for each direction viz. horizontal, vertical, $+45°$, $-45°$. Compute value of R for all four masks, the direction for which value of R is maximum, the point is more likely to contained on that direction.

(c) Edge Detection

Edge is boundary between the regions which have distinct intensity levels.

Various operators used for edge detection are:

(a) First Derivative:

First derivative operator responds whenever there is discontinuity in the intensity levels that is whenever there is transition from brighter intensity to darker intensity or vice versa.

(b) Second Derivative:

Second derivative operator gives positive sign on darker side and negative sign on brighter side. As second derivative operator is very-very sensitive to noise so we do not use second derivative operator for edge detection.

But we can use second derivative for extraction of secondary information such as we can use sign of second derivative to determine wether a point is on darker side or on brighter side of edge.

(c) Mask Use To Compute Image Gradient:

For computing image gradient two types of operators are used Perwitt Edge operator and Sobel edge operator.

a) Perwitt Edge Operator: In Perwitt edge operator two types of mask are used one for horizontal direction and other for vertical direction.

b) Sobel Edge Operator: In Sobel edge operator also there are two masks each for horizontal and vertical directions.

Sobel edge operator gives an averaging effect on image. So effect due to noise is taken care by sobel edge operator but not by perwitt operator.

(ii) Similarity base image segmentation.

There are mainly three types of similarity based segmentation techniques

1 Thresholding technique

In Thresholding technique we choose threshold in valley between two peaks, if we have multimodal histogram for an image we get multiple thresholds. Based on the value of intensity of pixel i.e greater or lesser than threshold we can determine pixel belongs to which region.

Various thresholding techniques are:

(a) Global Threshold Technique

(b) Dynamic or Adaptive Threshold Technique

(c) Optimal Threshold Technique

(d) Local Threshold Technique

2 Region Growing Technique

Basic idea behind region growing technique is we start from single point and try to find what other points can be included in the group. We select some points in image as seed point or starting points and then we look neighbourhood around seed points for similarity measures. Similarity measure can be intensity values. If pixels are similar then we will include them in a group.

Region Growing Technique is a recursive algorithm. For better results first we will use threshold technique to find seed points, then we apply region growing approach on original image.

3 Splitting And Merging

This is third type of Similarity Based Segmentation Technique. In this technique first of all we will divide images into various subimages and then merge subparts of images which are similar based on some similarity measure. This is an top down approach as it start from whole image and keep on dividing image into subparts and then this process continues recursively until no further splits or merges are possible.

1.2 Litreture Review on Satellite Image Segmentation

This section describes a brief review on different clustering techniques used for Satellite Image Segmentation, various supervised learning classifiers and technique used for evalution of optimal number of cluster.

1.2.1 Fuzzy Clustering Algorithms [3]

The basis of fuzzy set theory, established by Zadeh. Fuzzy set theory gives the idea of probability of belonging which is calculated by using membership function. After the introduction of fuzzy set theory, this technique have been used in many areas to solve complex and challenging problems. Due to feature of fuzzy logic that it is insensible to noise and can easily handle multidimensional feature in digital images it has been widely used for image analysis.

Traditionally, clustering algorithms divides the data into groups or clusters, where objects assigned to the same cluster are homogeneous according to some features, such as intensity or texture and where the objects classified to different regions are not similar. Nevertheless, there is usually no sharp boundary between clusters so the clasical hard clustering methods, such as the k-means, are not the most suitable for this kind of data. By contrast, within fuzzy logic, one of the most applied techniques are fuzzy clustering algorithms. These algorithms have considerable benefits than hard ones, because they could provide much information from the input image and they are not forced to classify each pixel of an image exclusively to one class. In this sense, fuzzy clustering algorithms allow objects to belong to multiple clusters, using a membership factor, that indicates with which grade is classified each object into each segmented cluster.

Different fuzzy clustering algorithms, particularly Fuzzy C-means and its variants, has been widley applied in the task of image segmentation. These algorithms classify the image into clusters, by iteratively minimizing a cost function in the feature domain that is dependent on the distance of the pixels to the clusters centers, also named centroids. Different Fuzzy clustering algorithms used in proposed algorithm are Fuzzy C Mean(FCM), Differential Evaluation Fuzzy C Mean(DeFCM) and Modified Differential Evolution Fuzzy C Mean(MODeFCM) algorithm.

1.2.2 Support Vector Machine (SVM) Ensemble Fuzzy Clustering (SVMeFC) [2]

Support Vector Machine (SVM) Ensemble Fuzzy Clustering (SVMeFC) algorithm uses all previously mentioned fuzzy clustering algorithms and SVM classifier as well. To train SVM classifier result of each fuzzy clustering algorithm is used, to do this some percentage of data are selected from each cluster, based on their distance from their respective cluster centers. These data points are considered as high confidence points. Then by using trained SVM classifier classes of the remaining points of image is determined. Then by using cluster-based similarity partition algorithm (CSPA) final clustering solution is generated.

In this newely developed SVMefc approach some of well known fuzzy clustering algorithms are used to get different clustering solutions. Fuzzy clustering algorithms used are such as Fuzzy C Mean,DeFCM MODeFCM. To get better result each clustering solution is used and given equal importance. To ensemble clustering solution from various fuzzy clustering algorithms CSPA is used.

The Steps of SVMeFC are:

1) Apply MoDEFC, GAFPSC, FCIDE, GAFC, and FCM separately with known K on the given data set to generate a set of solutions r_1, r_2, . . . ,r_N, where N is the number of fuzzy clustering methods. Each solution set consists of the corresponding cluster centers.

2) Repeat the following steps N times for each set of solutions r_i, i = 1, 2, . . .,N.

    (a) Select p no. of data points from each cluster, which are nearest to the respective cluster centers for each r_i.

(b) Train the SVM classifier by the selected training points.

(c) Predict the class labels for the remaining points (test points) using the trained SVM classifier.

(d) Combine the label vectors of training and testing points to obtain the label vector B_i, where i = 1, 2, . . .,N, for the complete data set, corresponding to the solution r_i.

3) Apply CSPA on the ensemble label vectors B_1, B_2,. . . , B_N to obtain the final clustering label vector.

The size of training data and test data depends upon percentage of points selected as high confidence points depending upon their distance from their respective cluster centers. If p that is number of points selected as training point are increased size of training data will also increases but it indicates low confidence of training set. But if p is decreased size of training data will also get decreased but it indicates high confidence of training set.

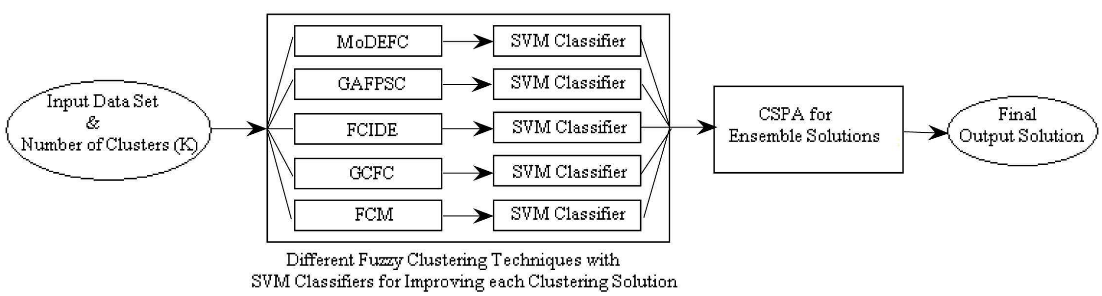


Figure 1.1: SVMeFC

1.2.3 Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation [8]

The main disadvantage of clustering algorithm is that we have to supply value of K, i.e we have to pass number of cluster as a parameter. In this paper intra cluster and inter cluster distance measures are used. Intra cluster distance is calculated as distance between a point and center of its respective cluster.

$$intra = \frac{1}{N}\sum_{i=1}^{K}\sum_{x\epsilon c}^{c}\|x - z_i\|^2 \tag{1.1}$$

where N denotes number of pixels in an image and K denotes number of clusters and z_i is the cluster center of i_{th} cluster.

The inter cluster is calculated by taking minimum of the distance between cluster centers as shown in next equation.

$$inter = min\left(\|z_i - z_j\|\right)^2 \tag{1.2}$$

where i=1,2,...K-1 and J=i+1.....K then we take ratio of these measures

$$validity = \frac{intra}{inter} \tag{1.3}$$

The process follwed is that an upper limit of number of clusters is taken that is K_{max} then validity measure is calculated for each value from 2 to K_{max} to determine which is best clustering solution. The process followed in this algorithm is that first of all the pixels of an image are present in one cluster then an iterative process is used where termination criteria is that number of clusters is equal to K_{max}, for splitting variance of cluster is calculated and the cluster whose variance is maximum is splitted into two. And then for this new number of clusters K means clustering algorithm is used to get clustering solution. After that to determine optimal value of k validity measure is calculated for each of them.
Variance of three components of cluster c_i is calculated as:

$$\sigma_{i,j}^2 = \frac{1}{N}\sum\left(x - z_{ij}\right)^2 \tag{1.4}$$

where i=1,2...K and j=1,2,3

1.2.4 Support Vector Machine [16, 17]

A support vector machine is a two class classifier that is it classifies data by using seperating hyperplane. After training of SVM using training data along with its class labels,support vector machine classifies new data that is test data and gives an hyperplane for test data.

The SVM algorithm work by finding the hyperplane which is at the largest possible minimum distance to the training samples. Two times of this distance is also known as margin. Since the goal of SVM is to maximize this margin of training samples as much as possible.

Samples closest to the separating hyperplane are the most difficult sample to classify, these samples are known as support vectors.

Support Vector Machine can also perform non linear classification by using Kernel functions. Kernel function works by mapping their input data to higher dimensions.

Given some training data a set of n points of the form

$$D = \{(x_i, y_i)\, x_i \epsilon R^p, y_i \epsilon \{-1, 1\}\}_{i=1}^{n} \tag{1.5}$$

where y_i is either 1 or -1. This y_i is the class label to which point x_i belongs that is either 1 or -1. Data in x_i are of P dimension. We want to find the maximum marginal hyperplane that divides the point having

$y_i = -1$ Any hyperplane can be written as the set of points x satisfying equation

$$w.x - b = 0 \tag{1.6}$$

where . is dot product of w and x and w is the vector normal to the hyperplane. The parameter $\frac{b}{\|w\|}$ determines the offset of the hyperplane from the origin along the normal vector w. If the given training data can be seprated linearly, two hyperplanes can be selected such that these hyperplanes seprates data such that no data point lies between them. Distance between these two hyperplanes is known as margin and then we try to maximize this margin. By the following equation we can describe these hyperplanes.

$$w.x - b = -1 \tag{1.7}$$

and

$$w.x - b = 1 \tag{1.8}$$

By using geometry, we find the distance between these two hyperplanes is $\frac{2}{\|w\|}$, so we want to minimize $\|w\|$

Following constraint has been added so that data point do not fall in between margin. For each i either

$$w.x - b \geq 1 \tag{1.9}$$

when x_i belong to first class or

$$w.x - b \leq -1 \tag{1.10}$$

when x_i belong to second class

1.3 Motivation

By analyzing recent research studies we can conclude that fuzzy clustering algorithms retains more information as compared to hard clustering algorithms and gives good results, but in each fuzzy clustering algorithm there is some drawbacks so none of fuzzy clustering algorithm gives good result for all applications. Thus to remove this drawback, in SVMeFC all these fuzzy clustering algorithms are integrated or ensembled to get good clustering solutions.

In SVMeFC result of fuzzy clustering techniques is used to train SVM classifier by taking certain points which are high confidence points. Then trained SVM classifier is used to classify remaining points to there class labels.

Since SVM is Non probabilistic hard binary decision approach. Probabilistic predictions are important in classification and decision making problems. While RVM produce probabilistic prediction. RVM introduce Zero mean Gaussian priors over every weight w_i and one more drawback associated with SVM is that number of support vectors increases linearly with size of training data due to which computaion complexity also increases. Thus instead of using SVM, RVM is better approach. The idea of proposed work is to segment image using RVM classifier and to evaluate optimal number of cluster.

1.4 Thesis Layout

Rest of the thesis is organised as follows —

Chapter 2: RVM based Satellite Image Segmentation In this chapter, we proposed a method to efficiently segment an satgellite image using RVM and as well application of Davies-Bouldin Index is also used to evaluate optimal number of cluster.

Chapter 3: Conclusion and Future Work This chapter provides the comparison of other classifiers such as SVM, KNN classifier with proposed algorithm using RVM to segment an image.

Chapter 2

RVM based Satellite Image Segmentation

In this chapter we discuss Satellite image segmentation using RVM where RVM is trained using result of fuzzy clustering algorithm by taking high confidence points which are near cluster centers. Advantage with RVM is that when size of training data increases its computational complexity do not increases unlike SVM in which number of support vectors increases linearly with size of training data. And other disadvantage with SVM is that in SVM predictions of classes of test data are not probabilistic while in RVM predictions are probabilistic. But the problem with the use of RVM is that it is slower than SVM.

2.1 Fuzzy C Mean (FCM) [4]

There are lot of fuzzy clustering algorithms but among them fuzzy c-means (FCM) algorithm is the most popular method used in image segmentation because it can handle ambiguity and can provide more information than hard segmentation methods such as K-means. FCM algorithm works well for noise free images. Disadvantage with FCM is that it is very sensitive to noise. So for images which contains noise FCM is not ideal algorithm for clustering.

The Fuzzy C-Means was introduced by Bezdek, it divides a finite collection of elements into c fuzzy sets by minimizing the following objective function:

$$J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} (U_{ij})^m \, d^2(x_j, v_i) \tag{2.1}$$

$$d^2(x_j, v_i) = (x_j - v_i)^T A (x_j - v_i) \tag{2.2}$$

The steps of FCM algorithm are:

1) Initalize membership U of X belonging to cluster i such that

$$\sum_{i=1}^{c} (u_{i,j}) = 1 \tag{2.3}$$

2) Compute the fuzzy centroid V for i=1,2,3.....c. using

$$V_i = \sum_{i=1}^{n} (U_{i,j})^m X_j \div \sum_{i=1}^{n} (u_{i,j})^m \tag{2.4}$$

3) Upadte the fuzzy membership U using

$$U_{i,j} = \left(1 \setminus d^2(X_j, V_i)\right)^{1 \setminus m-1} \setminus \sum_{i=1}^{c} \left(1 \setminus d^2(X_j, V_i)\right)^{1 \setminus m-1} \tag{2.5}$$

4) Repeat steps 2) and 3) untill the value of j is no longer decreasing.
   FCM algorithm converges towards solution that is minima of j depending on the intial value of U but different value of U gives different solutions for objective function j.

2.2 Differential Evolution Fuzzy Clustering Algorithm [5]

The basic steps of DEFCM are as follows:

a) Vector representation and Population Initialization

Each vector contains real number which represent K cluster centers. For d-dimensional data, size of vector is d K, where the first N positions represent the d dimensions of the first cluster centre, the next N positions represent those of the second cluster centre, and so on. In each vector K cluster centers are encoded. Initially they are initialized to K randomly choosen points taken from data set. For each vector in the population same process is repeated.

b) Fitness computation

With each vector a objective function or fitness function is associated. This fitness function is used to represent degree of goodness of the solution encoded in the vector. The fitness of each vector is computed using this equation:

$$J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} (U_{ij})^m \, d^2 (x_j, v_i) \tag{2.6}$$

c) Mutation

Any vector of the population at time step t which is also called as generation has d components as data is of d dimension, i.e.

$$\jmath (t) = [G_{i,1} (t) , G_{i,2} (t) ....., G_{i,d} (t)] \tag{2.7}$$

For each individual vector $\partial_k (t)$ that belongs to the current population, we use Differential Evolution for randomly choosen three other individuals, i.e., $\partial_m (t) , \partial_r (t) , \partial_j (t)$ from the same generation where k,j,i and m are all distinct. It then calculates the (component wise) difference, scales it by a scalar F (usually ϵ [0, 1]), and creates a trail offspring $\partial_m (t + 1)$ by adding the result to $\partial_m (t)$. Thus, for the n_{th} component of each vector

$$\partial_{k,d}(t+1) = \partial_{m,d}(t) + F\left(\partial_{r,d}(t) - \partial_{\jmath,d}(t)\right) \qquad (2.8)$$

d) Crossover

In order to increase the diversity of the parameter vectors, crossover is introduced. To this end, the trial vector:

$$\jmath(t+1) = \left[G_{i,1}(t+1), G_{i,2}(t+1) ....., G_{i,d}(t+1)\right] \qquad (2.9)$$

is formed.

e) Selection

To decide that computed vector should become a member of generation G + 1 or not, the trial vector $U_k(t+1)$ is compared to the target vector $\partial_k(t)$ using the greedy criterion. If vector $U_k(t+1)$ gives smaller cost function value than $\partial_k(t)$ then $U_k(t+1)$ is set to $\partial_k(t)$, otherwise, the old value $\partial_k(t)$ is retained.

f) Termination criteria

After executing processes of mutation, crossover and selection for a fixed number of iteration the best vector from the last generation is choosen which gives the solution of the clustering problem.

2.3 Modified Differential Evolution Fuzzy Clustering Algorithm [6]

In Modified Differential Evolution (MoDE), Differential Evolution algorithm is little bit modified in mutation phase. During mutation phase an different approach is used to push trial vector towards global optima quickly. In the mutation process, three vectors are there out of which one represent global best (GBest) and the other one represent local best (LBest).

$$V_i(t+1) = G_{GBest}(t) + \alpha(G_{LBest}(t) - G_j(t)) \tag{2.10}$$

The other phases except mutation such as crossover and selection are same as in original differential evolution algorithm. Depending on in each generation one of the process is selected, is computed as 1/[1 + exp((1/generation))]. One of the process uses mutation as in the original differential evolution other generates new mutant vector using equation 2.10. As the generation is inversely proportional to so as the number of generation increases value of decreases in the range [1, 0.5], due to which probability of using modified mutation reduces. Also, when modified mutation is used in the initial stage, local best(LBest) which is best vector in current population has more contributions for generating the mutant vector than in the later stage.

2.4 Relevance Vector Machine [13, 14]

In this paper relevance vector machine is explained. The problem of linear regression can be viewed as finding parameter w and offset c so that we can find class label for an unknown input.

$$y = w^T x + c \tag{2.11}$$

Offset c is generally included into w and if there is nonlinear mapping then basis functions such as kernel functions can be used.

$$y = w^T \phi(x) \tag{2.12}$$

where $\phi(x)$ is linear to non linear mapping. While calculating w from training data some noise ϵ_i is assumed. So the equation is

$$t_i = y_i + \epsilon_i = w^T \phi(x) + \epsilon_i \tag{2.13}$$

where ϵ_i is assumed to be samples from gaussian noise with mean 0 and variance σ^2. By looking at all the training samples at the same time so that the vector t represents all the individual training points t_i and the $N \times M$ design matrix θ is constructed such that the i_{th} row represents the vector $\phi(x_i)$, we have:

$$P\left(t|x_i, w, \sigma^2\right) = \left(2\Pi\sigma^2\right)^{\frac{-N}{2}} exp\left\{\frac{-1}{2\sigma^2}\|t - \theta w\|^2\right\} \tag{2.14}$$

When we try to calculate realtionship between x and y we try to minimize complexity and so we also have constraint growth of the weight w and to do this we define an explicit prior probability distribution on w. For smoother function we use Zero-mean Guassian prior over w.

So to define explicit prior probability distribution on w we used inverse variance α_i of each w_i. By looking at all th N points at the same time so that i_{th} element in the vector represents α_i, we have:

$$p(w|\alpha) = \prod_{i=0}^{N} N\left(0, \alpha_i^{-1}\right) \tag{2.15}$$

This means that there is an individual hyperparameter α_i associated with each weight, modifying the strength of the prior thereon.
The posterior probability over all the unknown parameters, given the data, is expressed as

$$P\left(w, \alpha, \sigma^2|t\right) \tag{2.16}$$

We are trying to find the w, and σ^2 which maximize this posterior probability. We can decompose the posterior:

$$P\left(w, \alpha, \sigma^2|t\right) = P\left(w|t, \alpha, \sigma^2\right) P\left(\alpha, \sigma^2|t\right) \tag{2.17}$$

We need to find hyperparameter α and β to evaluate m and $\sum$.

$$lnP(t|\alpha, \beta) = \frac{N}{2}ln\beta - E(t) - \frac{1}{2}ln\left|\sum\right| - \frac{N}{2}ln(2\Pi) + \frac{1}{2}\sum_{i=1}^{M} ln\alpha_i \tag{2.18}$$

2.5 Proposed Satellite Image Segmentation technique

In the proposed technique first of all for an input image we will evaluate optimal number of clusters using Davies-Bouldin index. The davies-Bouldin index is based on ratio of within cluster distance and between cluster distance. In other words it is based on the ratio of inter and intra cluster distance. The equation used is as follows:

$$intra = \frac{1}{N}\sum_{i=1}^{K}\sum_{x\epsilon c}^{c}\|x - z_i\|^2 \tag{2.19}$$

where N denotes number of pixels in an image and K is the number of clusters and z_i is the cluster center of respective cluster.

The inter cluster is calculated as the distance between cluster centers and take the minimum of them.

$$inter = min\left(\|z_i - z_j\|\right)^2 \tag{2.20}$$

where i=1,2,...K-1 and J=i+1.....K then we take ratio of these measures

$$validity = \frac{intra}{inter} \tag{2.21}$$

We choose an upper limit let K_{max} for the number cluster and then we calculate validity measure for each of them. The minimum value of Davies-Bouldin index gives optimal clustering and maximum value represent worst case.

After evaluating optimal number of clusters for image i.e K we will pass this image along with calculated K in fuzzy clustering algorithm such as FCM, DeFC, MoDEFC. Then by using different clustering solutions obtained from clustering algorithm RVM is trained. For this some percentage of points are selected as training points from each clusters based on their distance from cluster centres, these points can be considered as high confidence points and the remaining low confidence points of image are classified on the basis of trained RVM classifier.

Since clustering is unsupervised learning method, we are using result of unsupervised learning classifier to train supervised learning classifier that is RVM.

For this purpose we are choosing some percentage of points to train RVM classifier. As mentioned training of RVM is slower as compared to SVM. Relevance Vector Machine takes $O(N^3)$ time. Functional form of both RVM and SVM is identical but RVM does probabilistic classification. In RVM we need to maximize this equation:

$$lnP(t|\alpha,\beta) = \frac{N}{2}ln\beta - E(t) - \frac{1}{2}ln\left|\sum\right| - \frac{N}{2}ln(2\Pi) + \frac{1}{2}\sum_{i=1}^{M} ln\alpha_i \qquad (2.22)$$

with respect to hyperparameters α and β.

The original algorithm for RVM is for linear classifier. RVM belongs to the family of linear classifier. Despite this it does not mean that non linear data cannot be classified, non linear classifier can be created by using kernel trick. Kernel functions are used to classify non linear data. Kernel function projects data to higher dimension space so that it can be linearly classified. Since the classifier is a hyperplane in high dimension space it may be non linear in the original input space [15].

Different types of kernels are there which are used to project data in higher dimension space so that can be linearly classified.Some commonly used kernels are:

- Polynomial Homogeneous Kernel

$$K(x_i, x_j) = (x_i \cdot x_j)^d \qquad (2.23)$$

- Polynomial Inhomogeneous Kernel

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d \qquad (2.24)$$

As RVM is not multi class classifier. So to classify image using RVM tree structure is used. In tree structure first of all samples are classified into two classes as pixels of one cluster in one class and remaining other pixels in second class. And following the same procedure all points of an image are classified.

After feature extraction to reduce dimension principal component analysis is used. It is one of the most popular technique for dimension reduction. Other techniques for dimension reduction such as linear discriminant analysis (LDA), cannonical correlation analysis (CCA) can also be used. By computing covariance, mean and eigen values it reduces dimensions.

After classifying all points using trained RVM classifier Cluster based Similarity partition algorithm (CSPA) is used to ensemble solutions from different clustering algorithm. CSPA is used in situation where it is desired to find a single clustering solution from number of different clustering solutions. CSPA works by calculating that by how many clustering algorithm data points are placed in same cluster. Its main task to claculate similarity between two data points which is done by analyzing that number of clustering algorithm are placing these two data points in same cluster. More clustering algorithm placing these points in same cluster more similar are these two data points [18].

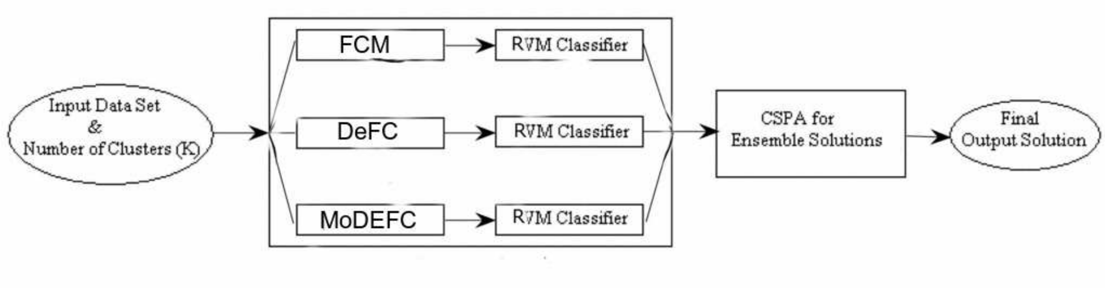
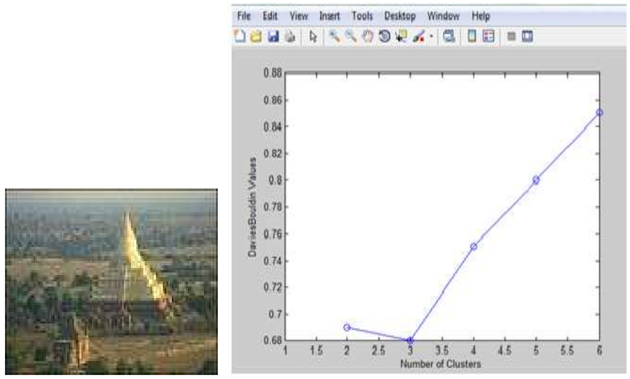


Figure 2.1: Image Segmentation using RVM

2.6 Results and Discussion

The proposed method is applied on various satellite image. All the computation is done in MATLAB on Intel n-series architecture with Core 2 Duo processer. Various metrics have been used for the comparison of results of proposed technique to other conventional techniques as discussed in chapter 1.

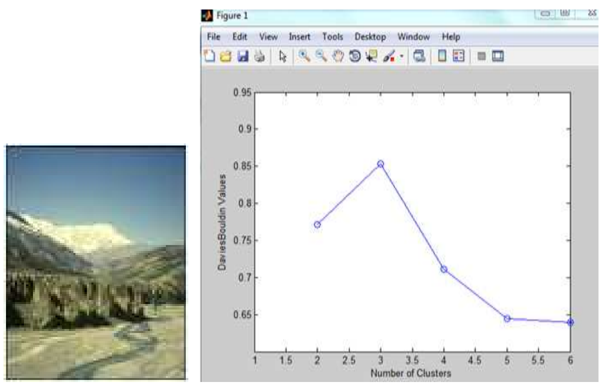
As discussed for the evaluation of optimal number of cluster Davies-Bouldin index is used, which is based on ratio of inter and intra cluster distance. Clustering for which Davies-Bouldin index is minimum gives optimal number of cluster.



(a) (b)

Figure 2.2: (a) Satellite-Image-1, (b) DBI1 value graph

As from Figure 2.2(b) we can clearly see that value of DBI validity measure is minimum for three clusters for image in Figure 2.2(a). For image shown in Figure 2.2(a) maximum number of cluster are taken as six and minimum value of DBI measure is for three clusters so optimal number of clusters for this image are three.

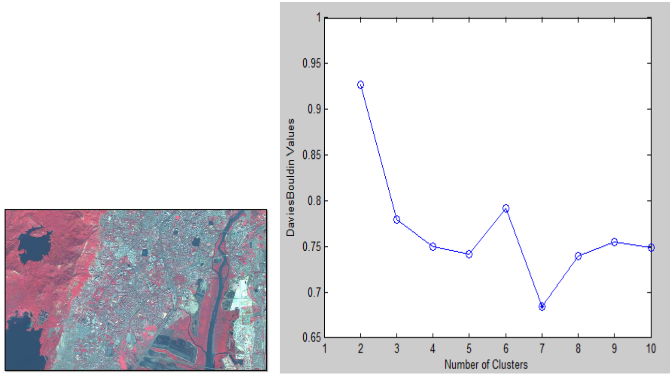


(a) (b)

Figure 2.3: (a) Satellite-Image-2, (b) DBI2 value graph

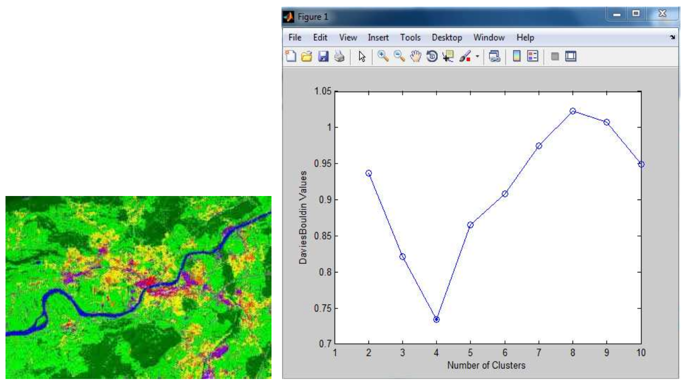
For images shown in Figure 2.3(a), 2.4(a) and 2.5(a) optimal number of clusters are six, seven and four respectively as calculated using value of Davies-Bouldin-Index validity measure.

After calculating optimal number of cluster for image using Davies-Bouldin index then various fuzzy clustering techniques are applied on satellite image. Fuzzy clustering algorithm used for this purpose are FCM, DeFC, MODeFC. As already



(a) (b)

Figure 2.4: (a) Satellite-Image-3, (b) DBI3 value graph



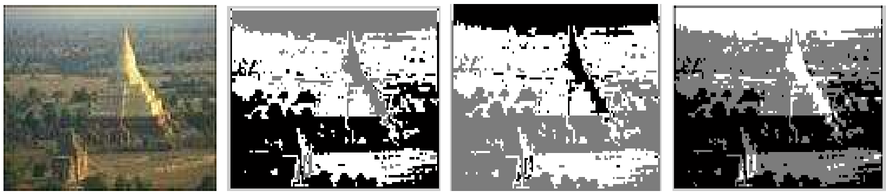
(a) (b)

Figure 2.5: (a) Satellite-Image-4, (b) DBI4 value graph

discussed that fuzzy clustering algorithms gives better result as compared to hard clustering algorithm as they uses uncertainity principle.

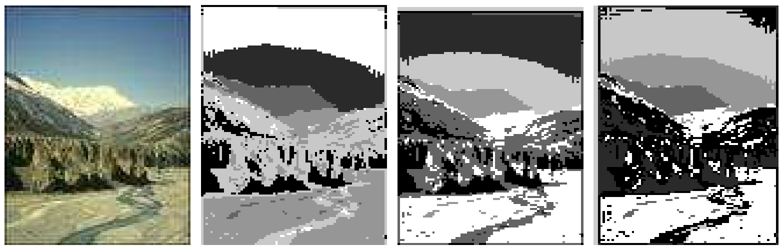
After calculating different clustering solutions using various fuzzy clustering algorithms RVM is trained by using certain percentage of points on the basis of their distance from their respective cluster centers these points are known as high confidence points. Then remaining pixels of the image are classified by using trained RVM classifier.

After applying Relevance Vector machine on image in Figure 2.6(a), 2.7(a), 2.8(a) and 2.9(a) result of segmentation is shown in Figure besides them using supervised learning RVM, KNN and SVM classifier.For the satellite-Image-1, Satellite-Image-2, Satellite-Image-3 and Satellite-Image-4 number of clusters taken



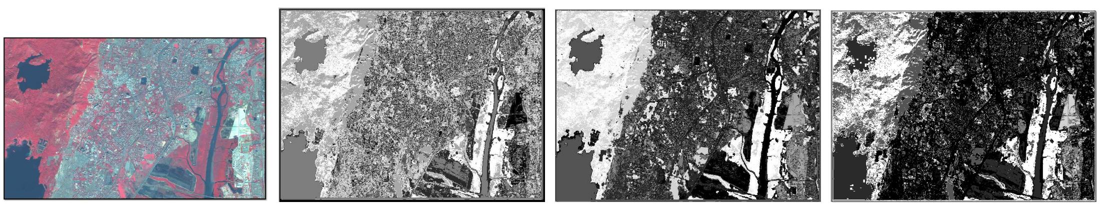
(a) (b) (c) (d)

Figure 2.6: (a) Satellite-Image-1, (b)RVM, (c)KNN, (d)SVM



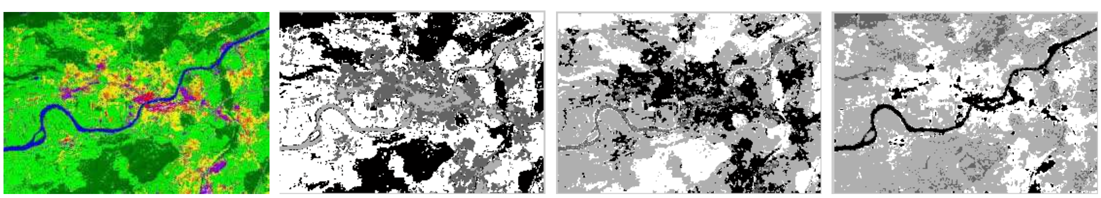
(a) (b) (c) (d)

Figure 2.7: (a) Satellite-Image-2, (b)RVM, (c)KNN, (d)SVM



(a) (b) (c) (d)

Figure 2.8: (a) Satellite-Image-3, (b)RVM, (c)KNN, (d)SVM



(a) (b) (c) (d)

Figure 2.9: (a) Satellite-Image-4, (b)RVM, (c)KNN, (d)SVM

are three, six, seven and four respectively. Four different satellite images are taken and after applying proposed technique result is shown. As we can see that proposed technique which uses RVM for classification gives good result. As we discussed that RVM does probabilistic prediction therefore it is better than SVM and KNN classifier. Result is compared on the basis of error rate and as compared to SVM and KNN, using RVM error rate is quite less. One drawback with RVM is that training of RVM is slow as compared to SVM and other supervised learning classifier. Result is also compared on the basis of time taken by this algorithm using different classifier that is RVM,SVM and KNN.

Table 2.1: Result of classifiers on Satellite-Image-1

Classifier	Error Rate (%)	Time (sec)
RVM	2.44	0.776
SVM	2.84	0.484
KNN	4.90	1.512

Table 2.2: Result of classifiers on Satellite-Image-2

Classifier	Error Rate (%)	Time (sec)
RVM	7.69	1.164
SVM	8.44	0.911
KNN	10.61	1.492

Table 2.3: Result of classifiers on Satellite-Image-3

Classifier	Error Rate (%)	Time (sec)
RVM	10.65	21.55
SVM	10.88	19.23
KNN	12.91	24.79

Table 2.4: Result of classifiers on Satellite-Image-4

Classifier	Error Rate (%)	Time (sec)
RVM	6.61	4.37
SVM	9.54	2.94
KNN	13.39	5.95

As from the result shown in table 2.1, 2.2, 2.3 and 2.4 it is clear that error rate for Satellite-Image-1, Satellite-Image-2, Satellite-Image-3 and Satellite-Image-4 using RVM is quite less as compared to KNN and SVM classifier. For four different satellite image different number of clusters are taken viz. three, six, seven and four respectively and then three different classifier RVM, SVM and KNN are applied and there result is compared.

2.7 Summary

In this chapter a satellite image segmentation technique by using RVM and various fuzzy clustering algorithm is explained. The proposed technique of satellite image segmentation produces good results in terms of quality than that of other available satellite image segmentation technique. This method is applied for various satellite images.

Chapter 3

Conclusions and Future Work

Among various image processing techniques such as image enhancement, image restoration, image segmentation is commonly used technique. This technique is used in various fields like medical, defense, industries. Despite image segmentation is very common but it is still very challenging task.

As satellite images covers very large land area and objects in satellite images are of non homogeneous type and satellite images are associated with large amount of imprecision and uncertainity so for satellite images hard clustering is not a good solution. Therefore various fuzzy clustering algorithms such as FCM, DeFCM are used for initial clustering of image pixels.

SVM is a supervised learning classifier. As in SVM with the increase in size of training data, number of support vectors increases linearly due to which time complexity of SVM also increases. While Relevance Vector Machine produces model with very few support vectors. One more drawback with SVM is that it does not uses probabilistic approach for prediction of classes while RVM does probabilistic classification.

KNN algorithm is very simple but main disadvantage with KNN algorithm is that it do not learn any thing from training data to predict class labels of new data it finds K closet neighbor of this new point from training set and assign it to most frequent class label among those K closet neighbor. Thats why KNN is also called as lazzy learner. And it also has to compute distance from training data

at each prediction which makes this algorithm slow if set of training example is large.
Thus RVM produces good results as compared to SVM and KNN classifier.

In this thesis an attempt have been made to propose an efficient satellite image segmentation technique. The proposed technique is explained in chapter 2. In proposed technique result of fuzzy clustering algorithms is used for training of RVM and then RVM is used to predict classes of remaining pixels of image. The proposed technique produces good results for satellite images.

Scope for Further Research

This thesis has opened several research directions which have scope of further investigation. This proposed algorithm can be extended to medical images. The computational complexity can be improve by using indexing for training of RVM.

Bibliography

[1] R. C. Gonzalez and R. E.Woods. Digital Image Processing. Addison Wesley,2nd edition, 1992.

[2] Saha, Indrajit, et al. "SVMeFC: SVM ensemble fuzzy clustering for satellite image segmentation." Geoscience and Remote Sensing Letters, IEEE 9.1 (2012): 52-55.

[3] Xie, Xuanli Lisa, and Gerardo Beni. "A validity measure for fuzzy clustering." IEEE Transactions on pattern analysis and machine intelligence 13.8 (1991): 841-847.

[4] Bezdek, James C., Robert Ehrlich, and William Full. "FCM: The fuzzy c means clustering algorithm." Computers and Geosciences 10.2 (1984): 191-203.

[5] Maulik, Ujjwal, and Indrajit Saha. "Differential Evolution Based Fuzzy Clustering Technique: Application to Satellite Image Segmentation." International Conference on Systemics, Cybernetics and Informatics. 2009.

[6] Maulik, Ujjwal, and Indrajit Saha. "Modified differential evolution based fuzzy clustering for pixel classification in remote sensing imagery." Pattern Recognition 42.9 (2009): 2135-2149.

[7] KARKKAINEN, ISMO, and PASI FRANTI. "Minimization of the value of Davies-Bouldin index." (2000).

[8] Ray, Siddheswar, and Rose H. Turi. "Determination of number of clusters in k-means clustering and application in colour image segmentation." Proceedings of the 4th international conference on advances in pattern recognition and digital techniques. 1999.

[9] Koonsanit, Kitti, Chuleerat Jaruskulchai, and Apisit Eiumnoh. "Determination of the Initialization Number of Clusters in K-means Clustering Application Using Co-Occurrence Statistics Techniques for Multispectral Satellite Imagery."

[10] Das, Swagatam, and Amit Konar. "Automatic image pixel clustering with an improved differential evolution." Applied Soft Computing 9.1 (2009): 226-236.

[11] Maulik, Ujjwal, and Indrajit Saha. "Automatic fuzzy clustering using modified differential evolution for image classification." Geoscience and Remote Sensing, IEEE Transactions on 48.9 (2010): 3503-3510.

[12] Joshi, Ameet, Shweta Bapna, and Sravanya Chunduri. ”Comparison Study of Different Pattern Classifiers.”

[13] Tzikas, Dimitris G., et al. ”A tutorial on relevance vector machines for regression and classification with applications.” EURASIP News Letter, 17, 2, 4 23 (2006).

[14] Tipping, Michael E. ”Sparse Bayesian learning and the relevance vector machine.” The journal of machine learning research 1 (2001): 211-244.

[15] Fletcher, Tristan. ”Relevance vector machines explained.” Practice (2008): 1-9.

[16] Cortes, Corinna, and Vladimir Vapnik. ”Support vector machine.” Machine learning 20.3 (1995): 273-297.

[17] Smola, Alex J., and Bernhard Schlkopf. ”A tutorial on support vector regression.” Statistics and computing 14.3 (2004): 199-222.

[18] Strehl, Alexander, and Joydeep Ghosh. ”Cluster ensembles—a knowledge reuse framework for combining multiple partitions.” The Journal of Machine Learning Research 3 (2003): 583-617.