Land Cover Classification Based Multiclass Support Vector Machine

Yin Yin Aye, Mya Thida Kyaw
University of Computer Studies, Mandalay
yinyinaye01.@.gmail.com

Abstract

Remote sensing data are attractive for deriving land cover information through the image classification. Most of the practical application involves multiclass classification, especially in remote sensing land cover classification. Support Vector Machine (SVMs) are based on statically learning methods and originally designed for the binary classification. A number of methods have been purposed to implement SVMs to produce multiclass classification. The main aim of this paper is to implement the supervised classification method for multispectral remote sensing images by using multiclass SVMs. This paper emphasizes on Directed Acyclic Graph (DAG), multiclass classification method on SVM. This system is implemented by JAVA Netbeans5.5 language.

1. Introduction

Land cover is one of the crucial elements for scientific research and real-life earth science applications. For many years, global, national, and regional managers and planners have recognized the importance of land cover for a variety of development of activities as it has also been used as a fundamental variable in several fields such as agriculture, environment, forestry, geology, and hydrology.

Due to the large scale proliferation of remote sensing data, they have become attractive sources of land cover information. Several classification algorithms have been developed and successfully implemented to produce land cover classification from multispectral data [1].

A new classification method based on novel statistical learning theory called the support vector machine has recently been applied to the problem of remote sensing data classification [2].

Support Vector Machines (SVMs) were originally designed for binary classification method, though, applications of binary classification are very limited. This technique is said to be independent of the dimensionality of the feature space as the main idea behind this classification technique is to separate the classes with a surface that maximize the margin between them, using the boundary pixels to create the decision surface. The data points that are closest to the hyper planes are termed as “Support Vectors” [3].

The Support Vector Machine (SVMs) is a theoretically superior machine learning methodology with great results in the classification of high dimensional datasets and has been found competitive with the best machine algorithms [2].

SVMs were tested and evaluated only as pixel based image classifiers with very good results. SVMs have often been found to provide better classification results that other widely used pattern recognition methods, such as maximum likelihood and neural network classifiers. Thus, SVMs are very attractive for the classification of remotely sensed data [4].

SVM method is more suitable for remote sensing image because that get with good result.

2. Image classification

Classification is one of the important stages in image processing. Remotely send data is used to assign corresponding levels with respect to groups with homogeneous characteristics, with the aim of discriminating multiple objects from each other within the image.

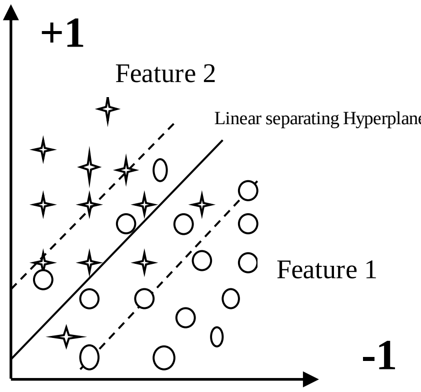
Classification can be executed on the base of spectral or spectrally defined features. It can be said that classification divides the feature space into several classes based on a decision rule[2].There are two classification methods for remote sensing image classification, supervised and unsupervised methods.

2.1. Supervised classification

In order to determine a decision rule for classification, it is necessary to know the spectral characteristic or features with respect to the population of each class. Some well known supervised classification methods are maximum likelihood, minimum-distance to mean, and parallelepiped method [2].

3. Support vector machine (SVM)

In this section, a brief introduction on how to construct an SVM is presented.



"Figure 1. Illustration of an SVM construction in a two dimensional feature space. Decision boundary with Maximum Mergin"

Consider a binary classification problem, where the given dataset is partitioned into two classes with a linear hyperplane separating them (figure 1). Assume that the training dataset consists of k training samples represented by $(x_1, y_1),\ldots\ldots,(x_k, k_k)$,where $x_i \in \mathfrak{R}^N$ is an N-dimensional data vector with each sample belonging to either of the two classes labeled as $y_i \in \{-1, +1\}$. The goal of SVMs is to find a linear decision defined by f(x) = w . x + b, where $w \in \mathfrak{R}^N$ determines the orientation of a discriminating hyperplane, and $b \in \mathfrak{R}$ is a bias. The hyperplanes for the two classes are, therefore represented by y_i (w .x + b)≥ 1.Sometimes, due to the noice or mixture of classes introduced during the selection of training data, variables $\xi_i > 0$, call slack variables, are used to account for the effects of misclassification. The hyperplanes for the two classes then become y_i (w .x + b)$\geq 1- \xi_i$. The optimal hyperplane (i.e.,f (x) = 0) is located were the margin between two classes of interest is maximized and the error is minimized. This can be achieved by solving the following constrained optimization problem,

Mnimize : $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{k}\xi_i$, (1)

Subject to : (w.x+b)$\geq 1\_\xi_i$, for i= 1,2,…,k.

The constant $0 < C < \infty$, called the penalty value or C value, is a regularization parameter. It defines the trade-off between the number of misclassifications in the training data and the minimization of the margin. In practice, the penalty value is selected by trial and error.

The constrained the optimization problem in (1) is solved by the method of Lagrange multipliers. The equivalent optimization problem becomes,

Maximize : $\sum_{i=1}^{k}\alpha - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\alpha_i\alpha_j y_i y_j (x_i - x_j)$,

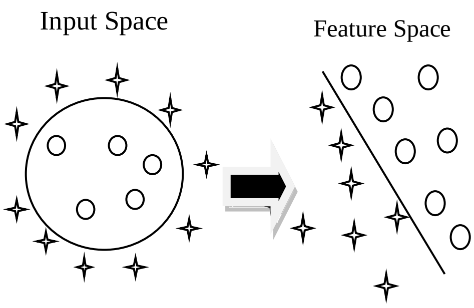
SubjectTo: $\sum_{i=1}^{k}\alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$, (2)

for i=1,2,…,k.

In (2), $\alpha_i \geq 0$ are the Lagrange multliers. The solution of the optimization problem given in (2) is obtained in terms of the Lagrange multipliers α_i. According to the Karush-Kuhn-Tucker(KKT) optimality condition, some of the multlipliers will be zero. The multipliers that have nonzero value are called the "support vector". The result from the optimizer, called an optimal solution, is the set $\alpha^0 = (\alpha_1^0, \ldots, \alpha_k^0)$. The value of w and b are calculated form $w^0 = \sum_{i=1}^{k} y_i \propto_i^0 x_i$ and $b^0 = \frac{1}{2}[w^0 x_{+1}^0 + w^0 x_{-1}^0]$, where x_{+1}^0 and x_{-1}^0 are the support vectors of class labels +1 and -1 respectively. The decision rule is then applied to classify the dataset into classes viz. +1 and -1, F(x) sign[$\sum_{Support\ vector} y_i \propto_i^0 (x_i, x) + b^0$] (3)

where sign(·) is the signum function. It returns +1 if the element is greater or equal to zero and -1 if it is less than zero.

Figure2 illustrates that two classes in the input space may not be separated by a linear separating hyperplane. However, the two classes are mapped by a nonlinear transformation function, a linear separating hyperplane can be found in the higher dimensional feature space.



"Figure 2. Mapping nonlinear data to a higher feature space"

Let a nonlinear transformation ø map the data to a higher dimensional space. Suppose there exists a function K , called a kernel function, such that,

$K(x_i,x_j) \equiv ø(x_i) . ø(x_j)$. (4)

A kernel function is substituted for the dot produt of the transformed vectors, and the explicit form of the transformation function ø is not necessarily known. Further, the use of the Kernel function is less computationally intensive.The optimization problem then becomes,

Maximize:$\sum_{i=1}^{k}\alpha - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}\alpha_i\alpha_j y_i y_j K(x_i - x_j)$

SubjectTo: $\sum_{i=1}^{k}\alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$, (5)

for i=1,2,…,k.
The decision becomes,
F(x)=sign[$\sum_{support\ vector} y_i \propto_i^0 K(x_i, x) + b^0$] (6) [1].

3.1. Characteristic of SVM

SVM has many desirable qualities that make it one of the most widely used classification Algorithms. Following is a summary of general characteristics of SVM:

1. The SVM learning problem can e formulated as a convex optimization problem, in which efficient algorithms are available to find the global minimum of the objective function.
2. SVM performs capacity control by maximizing the margin of the decision boundary. Nevertheless, the user must still provide other parameter such as the type of kernel function to use the cost function for introducing each slack variable.
3. SVM can be applied to categorical data by introducing dummy variables for each categorical attribute value present in the data.
4. The SVM formulation presented in this thesis is for binary class problems. Some of the methods available to extend SVM to multiclass problems are present [5].

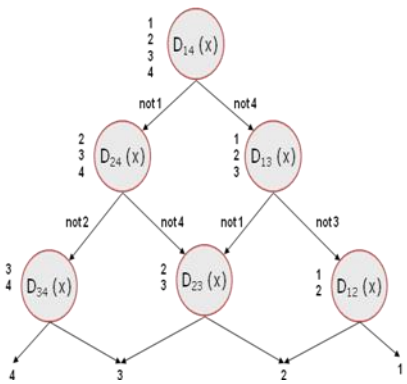
4. SVM for multiclass classification

Originally, SVMs were developed to perform binary classification. However, applications of binary classification are very limited especially in remote sensing land cover classification where most of the classification problems involve more than two classes. There are two types of approaches for multiclass SVM. One is by constructing and combining several binary classifiers. The other is by directly considering in one optimization formulation. Therefore, for multiclass SVM methods, they require to construct either several binary classifiers or a larger optimization problem [1]. There are:

1. One Against One (1A1)
2. One Against All (1AA)
3. Directed Acyclic Graph (DAG)

4.1. Classification based on Directed Acyclic Graph (DAG)

This method is based on the Decision Directed Acyclic Graph structure that has a tree-like structure. Similar to pairwise classification method and then create M (M-1)/2 binary classifiers for an M class classification. Each binary classifier is trained to distinguish two classes and forms a node in the graph structure. Nodes are organized in the form of a triangle with the single root node at the top and increasing subsequently in an increment of one node in each level until the last level that will have M nodes (see Figure 3).The DAG evaluates an input data starting at the root node and moves to the next level based on the output values. The binary classifier in the next level then evaluates the input data. The path traversed by data is called the evaluation path. The DAG method eliminates one class out from the list at each level. At the root node, all classes are in the list. Each node discriminates between the first class and the last class in the list. Each level gives the result in one class out of the two classes; the class that is not in favor of that level is eliminated from the list. The procedure is terminated when only one class remains in the list. Although, here the number of binary classifiers equals the number of classifiers required by the pairwise classification method, inputs are evaluated only M-1 times resulting in faster classification [1].



"Figure 3. The Directed Acyclic Graph SVM"

4.2. Multiclass objective function

Instead of creating many binary classifiers to determine the class labels, this method attempts to directly solve a multiclass problem. This is achieved by modifying the binary class objective function and adding a constraint to it for every class. The modified objective function allows simultaneous computation of multiclass classification and is given by

$\text{Min}_{w,b,\xi}$ [$\frac{1}{2}\sum_{i=1}^{M}\|W\|^2 + C\sum_{i=1}^{k}\sum_{r\neq y_i}\xi_i^r$]

Subject to the constraints,

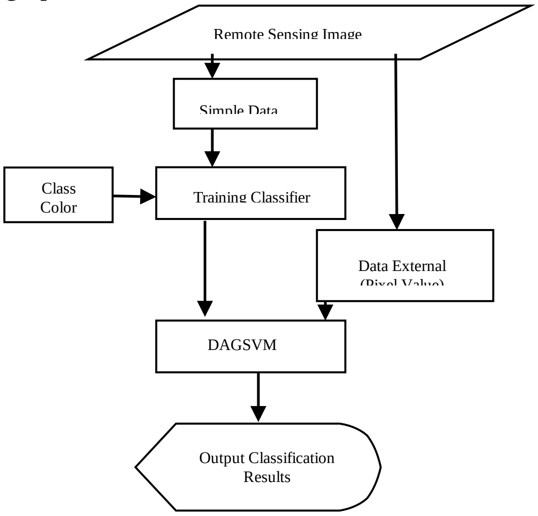
$w_{y_i}.x_i + b_{y_i} \geq w_r.x_i + b_r + 2 - \xi_i^r$ for

And, $\xi_i^r \geq 0$ for $i = 1, \dots, k$

Where y_i ϵ{1,.....,M} are the multiclass levels of the data vectors and rϵ{1,....,M}\ y_i are multiclass labels excluding y_i.

5. Overview of the system

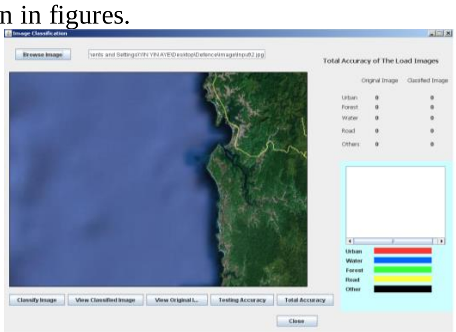
In this system, Remote Sensing Images are used as the input. Remote Sensing Images get from by using http:// Googlearch.com. This system trained with eight images and any remote sensing image can be train but one process one image can be use. This system consists of two parts. There are training and testing. In training, select the simple data from Remote Sensing Image. In this step, get the class label and multispectral features for each simple data. And then trained with DAGSVM, get the decision function. In this phase, support vector to determine the margin from previous phase. By using these margins can separate the multiclass of the image. In testing phase, make classification on the test data by using above model. Finally, get the classified output graphical view.



"Figure 4. System Flow"

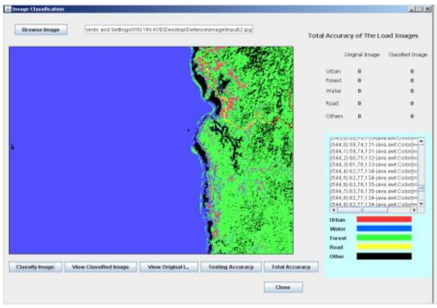
6. Implementation

When the Remote Sensing Image classified, get five outputs are urban, water, forest, road and other. Those are present with represented colors. Examples of input images, result image and accuracy are shown in figures.



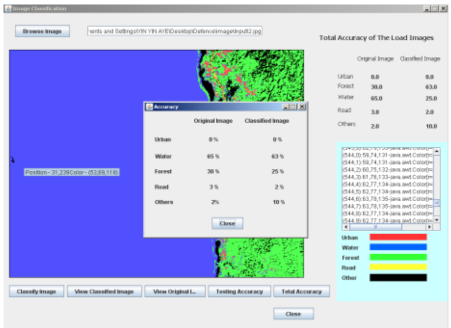
"Figure 5. Example of input Image1"

In the presenting original image consist of urban 0%, water65%, forest30%, road3% and other are 2%.

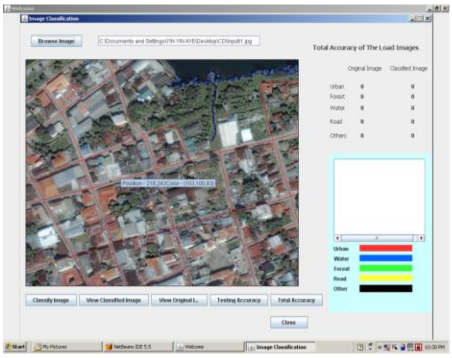


"Figure 6. Example of result Image1"

The size of the image is limited as 900*730.In the result image consist of urban 0%, water 63%, forest25%, road2% and other are 10%.

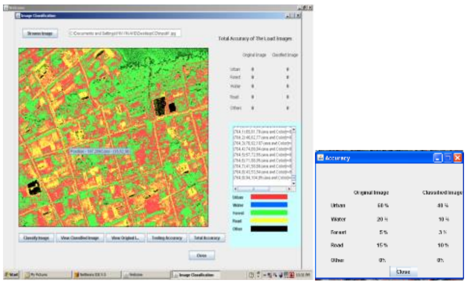


"Figure 7. Accuracy of the Result1"

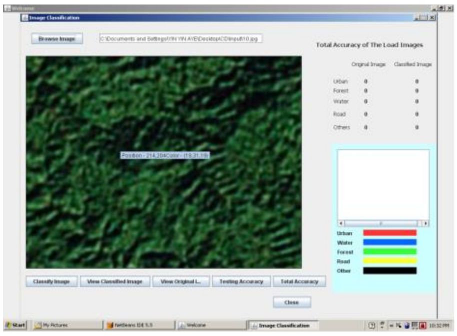


"Figure 8. Example of input Image2"

In the presenting original image consist of urban 60%, water20%, forest5%, road15% and other are 0%. In the result image consist of urban 40%, water 10%, forest3%, road10% and other are 0%.

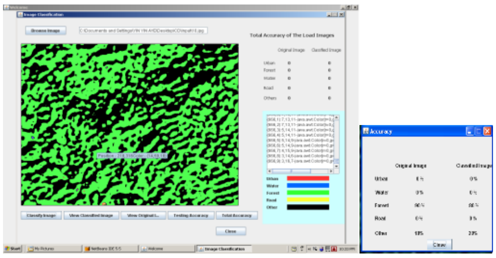


“Figure 9. Accuracy of the Result2”

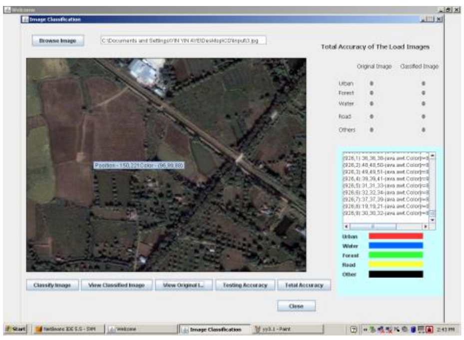


“Figure 10. Example of input Image3”

In the presenting original image consist of urban 0%, water0%, forest90%, road0% and other are 10%. In the result image consist of urban 0%, water 0%, forest80%, road0% and other are 20%.

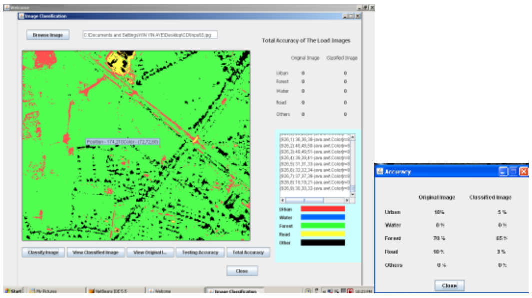


“Figure 11. Accuracy of the Result3”

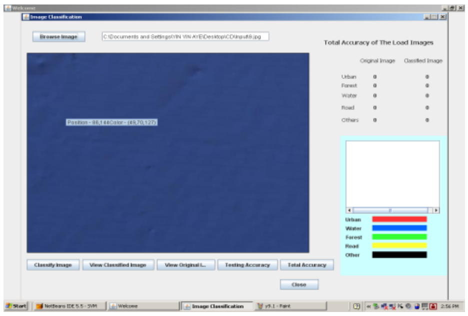


“Figure 12 . Example of input Image4”

In the presenting original image consist of urban 10%, water0%, forest70%, road10% and other are 10%. In the result image consist of urban 5%, water 0%, forest65%, road3% and other are 17%.

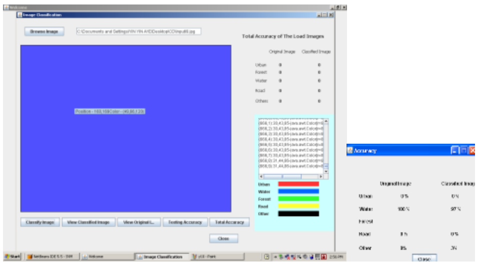


“Figure13. Accuracy of the Result4”

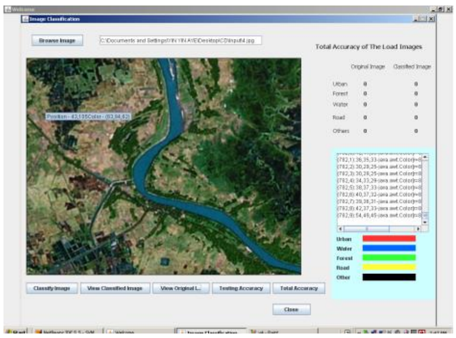


“Figure 14 . Example of input Image5”

In the presenting original image consist of urban 0%, water100%, forest0%, road0% and other are 0%. In the result image consist of urban 0%, water97%, forest0%, road0% and other are3%.

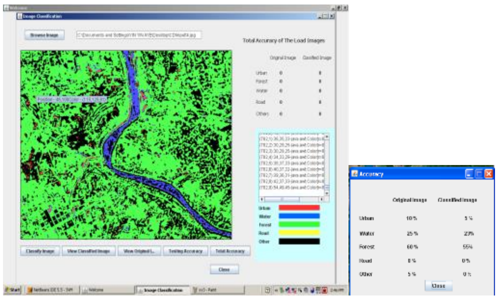


“Figure 15. Accuracy of the Result5”

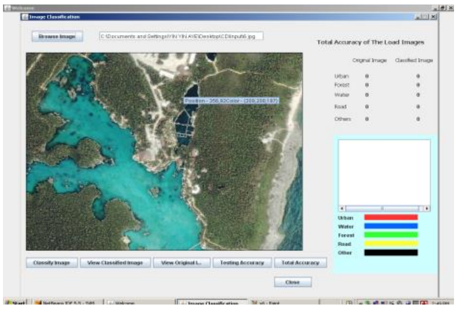


“Figure 16. Example of input Image6”

In the presenting original image consist of urban 10%, water25%, forest60%, road0% and other are 5%. In the result image consist of urban 5%, water 23%, forest65%, road0% and other are 7%.

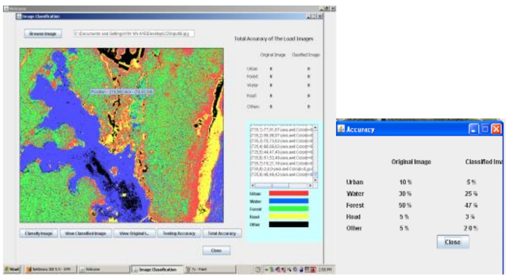


"Figure17. Accuracy of the Result6"



"Figure 18. Example of input Image7"

In the presenting original image consist of urban 10%, water30%, forest50%, road5% and other are 5%. In the result image consist of urban 5%, water25%, forest47%, road3% and other are20%.



"Figure 19. Accuracy of the Result7"

7. Conclusion

Classification of surface features in satellite imagery is one of the most important applications of remote sensing. Many research works have been made on classification by using neural networks and statistical methods. The classification problem involved the identification of five land cover types (urban, water, forest, road, other). A total of eight images were trained for all five classes. Future work is required to study the use of SVM approach with other type of multiclass classification method.

8. References

[1] P.W,Pramod.K.Varshney,Manoj K.Arora, Support Vector Machines For Hyperspectral Classification.

[2] T. Z. Ohn, M. Moe and Z. Htet. Multiclass Support Vector Machines for Remote Sensing Image Classification *at University of Computer Studies, Yangon*.

[3] M.Pal, Kurukshetra, Multiclass Approaches for Support Vector Machines Based Land Cover Classification.

[4] Angelos Tzotsos,Laboratory of Sensing,Technical *University of Athens,Greece*-tzotsos@gmail.com, Support Vector Machines approaches for object based image analysis.

[5] J. Wetson, *Data mining*, Second Edition (Support Vector Machine).