Geophysical Research Abstracts, Vol. 7, 04471, 2005
SRef-ID: 1607-7962/gra/EGU05-A-04471

Land cover classification by remote sensing: Potential of support vector machine classifiers

Ajay Mathur and Giles M. Foody

School of Geography

University of Southampton

Southampton

SO17 1BJ

United Kingdom

A.MATHUR@soton.ac.uk, G.M.Foody@soton.ac.uk

Land cover affects our climate by influencing energy, water and gas exchanges with the atmosphere and acts both as source and sink for biogeochemical cycles. Accurate information on land cover is, therefore, required to aid the understanding and management of the environment. Supervised classification is one of the widely used techniques in remote sensing to map land cover classes.

The accuracy of supervised classification is dependent to a large extent on the input training data. In general, the analyst aims to capture a large training set to fully describe the classes spectrally, often with the requirements of a conventional statistical classifier in mind. However, it is not always necessary to provide a complete description of the classes if using support vector machine (SVM) as the classifier.

A key attraction of the SVM based approach to classification is that it seeks to fit an optimal hyper-plane between the classes. Since the SVM uses only the training

samples that lie at the edge of the class distributions in feature space (support vectors) it may require only a small training sample. It may, therefore, be possible to use only "border" training data in training the classifier. Thus the classifier banks not only on the size of input training data but depends more on its location in the feature space.

The aim of the paper is to highlight the effect of classifier choice with particular regard to classification accuracy. The classifiers investigated were the widely used maximum-likelihood classifier (MLC), feed forward neural networks (NN), decision trees (DT) and support vector machines (SVM). Indian Remote Sensing Satellite (IRS-1D) with a spatial resolution of approximately 24 m acquired by LISS-III sensor, date of pass 22^{nd} September, 2003 path/row of 93/49 and 93/50 were used. The data set comprised three bands red, near-infrared and middle-infrared and corresponds to electromagnetic spectrum range of 0.62-0.68 μm, 0.77-0.86 μm and 1.55-1.75 μm respectively.

Three agricultural classes namely; cotton, rice (basmati), rice (local) were mainly grown in the study area and were the focus of this study. Classes built-up and sand which are difficult to segregate using conventional classifiers were also included in the study. A stratified random sample of 90 pixels per-class was derived for each class for training the classifiers and an equal number for testing.

The quality of water in the study area is not suitable for agriculture. To supplement appropriate water for agriculture, a number of canals criss-cross the area. A consequence of the use of canal water is that ground water is not being drawn for agriculture and hence the water table is rising. This has also resulted in waterlogging at many places in the study area. The waterlogging adversely affects the growth of plants as soil gets affected with salinity accompanied with the increased risk of insect, pests and pathogen attacks on the crops. This has adversely affected the crops, particularly cotton as the other major crop, rice can withstand water. The cotton has, therefore, often been replaced by rice, a water intensive crop. The diversion from cotton to rice requires more canal water, which is having multiplicative effect on the waterlogging problem. A number of anti-waterlogging measures have been undertaken in the area and the relevant agencies are interested to have an accurate land cover map over years to tackle the problem of waterlogging and maintain water balance in the area.

The SVM produced the largest accuracy (92 %) of the four classifiers. Furthermore, it used only 215 training pixels as support vectors out of the possible 450 used by other classifiers for training. Since, the number of support vectors used was a fraction of the total input training data, there appears to be a potential to use small training set size with SVM classifier without any negative impact on classification accuracy. This indicates that training data are not equally important and the importance of each training data sample in establishing decision boundaries is dependent on its location

in feature space, with border training data more important than those lying away from border in feature space.

The study concludes that SVM classifier, have the potential of using a small training data as compared to other conventional classifiers like DA, DT and ANN. The higher generalization ability of SVM with fast implementation makes it a strong candidate for classifying remotely sensed data. The accurate land cover maps will help the geo-scientists in tackling the problem of waterlogging in the study area.