A New Tree-Based Classifier for Satellite Images

Reshu Agarwal
Acadia University, Canada

Pritam Ranjan
Acadia University, Canada

INTRODUCTION

Remote sensing methods have extensively been applied in many areas, for instance, hydrological modeling, wildlife habitat modeling, forest degradation monitoring, wetland biodiversity conservation and disaster management. Different types of sensors mounted on a satellite generate images with different spatial, spectral, radiometric and temporal resolutions. Due to synoptic view and repetitive coverage, remote sensing is a fast emerging source of large datasets on natural resources in spatial format.

One of the most important products of a raw satellite image is the classified map which labels the image pixels in meaningful classes. Though several parametric and non-parametric classifiers have been developed so far, reliable prediction of pixel-wise class labels still remains a challenge. The inaccuracy and uncertainty in prediction can often be attributed to the complexity of study area terrain, sensor characteristics, spectral mixing and size of the training data.

In this chapter, we are promoting a new classifier called *mBACT* – a multiclass generalization of Bayesian Additive Classification Tree (BACT) for classifying satellite images. BACT is based on the *ensemble of trees* model called Bayesian Additive Regression Tree (BART) proposed by Chipman, George & McCulloch (2010), as a flexible regression model. Chipman et al. (2010) also developed a binary classifier called BART-probit for a drug discovery application. Independently, Zhang & Hardle (2010) used BART to develop BACT for classifying binary data in credit risk modeling setup. We compare the performance of mBACT with several state-of-the-art classifiers in remote sensing literature (i.e., support vector machine (SVM) and classification and regression tree (CART)) for predicting pixel-wise class labels of a satellite image. The data considered in this study is a portion of a LANDSAT 5 TM image (with six reflectance bands) covering the town of Kentville, Nova Scotia, Canada (see Agarwal, Ranjan & Chipman, 2013 for a detailed case study).

BACKGROUND

Classified maps play an important role in numerous remote sensing data applications, for example, land-cover change, forest degradation, hydrological modeling, wildlife habitat modeling, and biodiversity conservation. One of the most important products of a raw image is the classified map which labels the image pixels in meaningful classes. Several classifiers have been developed (e.g., Franklin, Peddle, Dechka & Stenhouse, 2002; Pal & Mather, 2003; Gallego, 2004) and implemented worldwide in software packages (e.g., ERDAS IMAGINE, ENVI, IDRISI and ArcGIS) for classifying satellite images. However, accurate prediction of pixel-wise class labels is still a challenge (Blinn, 2005; Song, Duan & Jiang, 2012).

Among various classification methods, Maximum Likelihood (ML) classifier is the most widely used classifier because of its simplicity and availability in image processing softwares (Peddle,

DOI: 10.4018/978-1-4666-5202-6.ch003

1993). ML classifier is based on a parametric model that assumes normally distributed data, which is often violated in complex landscape satellite images (Lu & Weng, 2007). Non-parametric classifiers do not require stringent model assumptions like normality and gained much popularity. For instance, the classifiers based on k-nearest neighbor (k-NN), artificial neural network (ANN), decision trees and support vector machines (SVM) have shown better performance as compared to ML classifiers (Zhang & Wang, 2003; Bazi & Melgani, 2006; Li, Crawford & Jinwen, 2010; Atkinson & Naser, 2010). Comparison of the classifiers has been an active research area in machine learning. For example, Sudha & Bhavani (2012) concluded that SVM is better classifier than k-NN; and Song et al. (2012) demonstrated that SVM and ANN are comparable; however, SVM often performs slightly better than ANN.

Decision tree based classifiers became very popular in machine learning literature after classification and regression tree (CART) was introduced by Breiman, Friedman, Olshen & Stone (1984). In remote sensing applications, CART has extensively been used for the classification of multispectral and hyper-spectral images (Yang et al., 2003). Refinements over CART (e.g., bagging, boosting and random forest) have also been used in remote sensing for more accurate class label identification (e.g., Lawrence, Bunn, Powell & Zambon, 2004). CART predictions are based on one tree, whereas BART (the underlying base model of the new classifier mBACT) uses an ensemble of trees-based model for predicting the class labels. As a result, mBACT leads to more accurate prediction of class labels.

METHODOLOGY

This section starts with brief overviews of CART, BART and BACT. Subsequently, we outline the new methodology of mBACT. See Brieman et al. (1984) for details on CART, Chipman et al. (2010) for BART and BART-probit, Zhang & Hardle (2010) for BACT, and Agarwal et al. (2013) for mBACT.

Classification and Regression Tree (CART)

Classification trees have gained much popularity in machine learning literature since CART was developed by Brieman et al. (1984). CART methodology has also been implemented in R software (R core development team, 2012) as a library called *rpart* (Therneau, Atkinson & Ripley, 2013). In the context of image classification, the main idea is to come up with a decision tree that partitions the image through recursive partitioning into homogeneous regions. We only discuss binary trees, as the decision trees with higher order splits can be obtained by iterative binary splits in binary trees.

Suppose the data consists of one response variable $y \in \{1, \ldots, n\}$ with n classes and p predictor variables denoted by $x = (x_1, \ldots, x_p)$. Let N be the total number of training data points. Then, the construction of the decision tree starts with assigning the entire training data in one group called the root node. This node is now split into two nodes via one of the p predictors. For instance, x_i, can be used to split the entire data into two subgroups or nodes $\{x: x_i \leq a\}$ and $\{x: x_i > a\}$.The two nodes are then further split using a value of another (or the same) predictor variable. The splitting process continues until a full tree is generated. Techniques like cross-validation are often used to prune the light branches (nodes with very few data points) which prevents over-fitting. Finally, each terminal (or leaf) node is assigned a class label from $\{1, \ldots, n\}$

For every node, the best splitting variable and split point combination $\{X_i, a\}$ is chosen by minimizing the impurity (or equivalently, by maximizing the homogeneity) among the predicted class labels within each partition. A few popular impurity indices are Gini index, entropy and misclassification error.