Using prior probabilities in decision-tree classification of remotely sensed data

D.K. McIver, M.A. Friedl*

Department of Geography, Center for Remote Sensing, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA

Received 27 November 2000; received in revised form 8 December 2001; accepted 8 December 2001

Abstract

Land cover and vegetation classification systems are generally designed for ecological or land use applications that are independent of remote sensing considerations. As a result, the classes of interest are often poorly separable in the feature space provided by remotely sensed data. In many cases, ancillary data sources can provide useful information to help distinguish between inseparable classes. However, methods for including ancillary data sources, such as the use of prior probabilities in maximum likelihood classification, are often problematic in practice. This paper presents a method for incorporating prior probabilities in remote-sensing-based land cover classification using a supervised decision-tree classification algorithm. The method allows robust probabilities of class membership to be estimated from nonparametric supervised classification algorithms using a technique known as boosting. By using this approach in association with Bayes’ rule, poorly separable classes can be distinguished based on ancillary information. The method does not penalize rare classes and can incorporate incomplete or imperfect information using a confidence parameter that weights the influence of ancillary information relative to its quality. Assessments of the methodology using both Landsat TM and AVHRR data show that it successfully improves land cover classification results. The method is shown to be especially useful for improving discrimination between agriculture and natural vegetation in coarse-resolution land cover maps. © 2002 Elsevier Science Inc. All rights reserved.

1. Introduction

Information from ancillary data sources has been widely shown to aid discrimination of classes that are difficult to classify using remotely sensed data (Brown, Loveland, Merchant, Reed, & Ohlen, 1993; Foody & Hill, 1996; Strahler, 1980). Such information is particularly useful in large area land cover mapping problems that involve cover types characterized by high levels of within-class variability (Loveland et al., 1999). One means of incorporating ancillary information is by using prior probabilities of class membership. Prior probabilities can improve classification results by helping to resolve confusion among classes that are poorly separable, and by reducing bias when the training sample is not representative of the population being classified.

Until recently maximum likelihood classification has been the most common method used for supervised classification of remotely sensed data (Richards, 1993). This methodology assumes that the probability distributions for the input classes possess a multivariate normal form. Increasingly, nonparametric classification algorithms are being used, which make no assumptions regarding the distribution of the data being classified (Carpenter et al., 1999; Foody, 1997; Friedl, Brodley, & Strahler, 1999). The use of such techniques has been motivated by the increased accuracy and flexibility that these algorithms provide. In particular, nonparametric classification methods have proven to be especially useful for continental and global-scale land cover mapping, where the classes of interest tend to be multimodal and possess substantial within-class variance (Friedl et al., 2000; Gopal, Woodcock, & Strahler, 1999; Hansen, DeFries, Townshend, & Sohlberg, 2000).

In this paper, we expand upon previous work using nonparametric classification algorithms for remote sensing applications (McIver & Friedl, 2001). Specifically, we explore the use of prior probabilities with decision-tree
classification algorithms. To accomplish this objective, the paper has four main components. First, a methodology to incorporate prior probabilities in nonparametric classification algorithms is presented. This methodology includes an approach for controlling the influence of prior probabilities depending on their quality. Second, we illustrate several issues regarding the influence of prior probabilities and training sample properties on classification outcomes from nonparametric classifiers. Third, we present an assessment of the methodology using Landsat Thematic Mapper (TM) data. This assessment uses field sites and prior probabilities derived from terrain data for an area characterized by strong topographic control over vegetation cover. Finally, the technique is applied to a coarse resolution land cover classification problem using Advanced Very High Resolution Radiometer (AVHRR) data. In this last case, prior probabilities estimated from data related to the global distribution of agricultural land use intensity are used to resolve confusion between natural vegetation and agriculture at continental scales.

2. Methodology

2.1. Boosting

Maximum likelihood classification has been widely used in remote sensing to estimate per-pixel conditional probabilities of class membership. While some nonparametric classification algorithms (e.g., decision trees) can also provide such estimates, they are typically quite approximate (Breiman, Friedman, Olshen, & Stone, 1984; Friedman et al., 2000). The approach used here is based on recent theory that explains the behavior of a machine learning algorithm known as boosting using statistical theory. This explanation provides a robust method to estimate conditional probabilities of class membership from nonparametric classification algorithms.

Boosting is one of numerous ensemble classification methods that are widely used in the machine learning research community (Freund, 1995). Ensemble methods are used in conjunction with a base classification algorithm (e.g., a decision tree or neural network) and have been widely shown to improve the performance of many classification algorithms. Specifically, boosting improves classification accuracies (Dietterich, 2000; Quinlan, 1996a), is resistant to overfitting (Bauer & Kohavi, 1999; Schapire, Freund, Bartlett, & Lee, 1998), and has been shown to be effective with remotely sensed data (Friedl et al., 1999; McIver & Friedl, 2001). While boosting is most commonly associated with decision trees (Breiman, 1998; Quinlan, 1996a), it has also been applied with other supervised classification algorithms such as artificial neural networks (e.g., Opitz & Maclin, 1999; Quinlan, 1996b). As a result, the method described in this work is not unique to decision trees and can be used with a variety of classification algorithms.

Boosting operates by iteratively estimating multiple classifiers using a base algorithm while varying the training sample. At each iteration, the training sample is modified to focus attention on examples that were difficult to classify correctly in the previous iteration. Modifications to the training sample are performed by maintaining a weight for each training sample that is adjusted each iteration. A weighted voting scheme is then used to select the predicted class. Friedl et al. (1999) and McIver and Friedl (2001) provide discussions of boosting for remote sensing applications. For a complete description of boosting, see Freund and Schapire (1997) or Friedman et al. (2000).

Recent theoretical work regarding boosting has shown it to be a form of additive logistic regression (Collins, Schapire, & Singer, 2000; Friedman et al., 2000). As a result, conditional probabilities of class membership can be obtained along with class predictions. From this perspective, boosting can be described as an additive model (Hastie & Tibshirani, 1999), where at each iteration the base learning algorithm is fit to the classification residuals (i.e., errors) from the previous iterations (Friedman et al., 2000). That is, the residuals from iteration \( n \) are used to modify the training sample used at iteration \( n+1 \) to emphasize previously misclassified examples. As Friedman et al. (2000) show, this interpretation allows conditional probabilities of class membership to be computed for each pixel.

A number of different boosting algorithms have been developed (Friedman et al., 2000; Schapire & Freund, 1999; Schapire et al., 1998). The algorithm used in this research, AdaBoost.M1 (Freund & Schapire, 1997), is the simplest multiclass boosting method. For this work, AdaBoost.M1 was implemented using the C4.5 decision tree (Quinlan, 1993), which is widely used in the machine learning community. The approach used throughout this paper is to obtain probabilities of class membership from boosting using 10 iterations of the base algorithm (C4.5). Previous research has shown that 10 or more iterations produces accurate probability estimates (McIver & Friedl, 2001). Outside information, which is expected to improve the classification, is then incorporated using Bayes’ rule in association with confidence estimates as described in Section 2.2.

2.2. Controlling the influence of prior probabilities

A common problem when using prior probabilities with supervised classification algorithms is that they can bias the posterior probability of class \( k \) given observation \( A \) (i.e., \( P(k|A) \)), towards the result predicted by the ancillary information (Strahler, 1980). This bias can be particularly problematic when there are errors or uncertainty in the information used to prescribe the prior probabilities. A
solution proposed by Chen, Ibrahim, and Yianoutsos (1999) (for use with Bayesian logistic regression models) provides an approach for addressing this problem.

In this approach, the ancillary information provides a likelihood estimate for the membership of each pixel in a given class \(j\) \(L_j\). \(L_j\) is then conditioned using a subjective confidence parameter \(c\) that ranges from 0.0 to 1.0. The conditioned estimate \(L_j^c\) can then be used as the prior probability in Bayes’ rule. As \(c\) is varied from 1 to 0, the effect of the likelihood estimate decreases until, when \(c\) equals 0, no information from the likelihood estimate is incorporated. Thus, the parameter \(c\) controls the influence of \(L_j\) on the predictions, based upon the user’s confidence in \(L_j\).

While the choice of a value for \(c\) is subjective, we show below that a simple heuristic based on objective criteria can be used to prescribe a value for \(c\) in many situations.

3. A simple example

In the following example, sample data were randomly drawn from two overlapping bivariate \((X,Y)\) normal distributions (Fig. 1). Each distribution possessed the same mean in the \(Y\) dimension and equal variances and covariances in both dimensions. They are differentiated only by their means in the \(X\) dimension. Probability density functions for each distribution exhibit considerable overlap in the \(X\) dimension and are shown in Fig. 1a. The training data were generated using a random sample from each distribution. Separate test data were then created using a systematic sample of the entire feature space.

3.1. Interaction between prior probabilities and class separability

To demonstrate the effect of prior probabilities on classification results, the training sample used for this exercise contained an equal number of examples from each class. The prior probability of Class 2, \(P(2)\), was then varied from 0.3 to 0.8. Changes in the proportion of predictions in the test data for each class arising as a result of varying \(P(2)\) are shown in Fig. 1b. As we would expect, the proportion of cases predicted to belong to Class 2 increases linearly with \(P(2)\).

A key point to note is that only cases that lie in the overlapping region of the two class distributions are affected.
by the prior probabilities. Cases in separable regions (roughly $X < 2$ and $X > 4$), on the other hand, are correctly classified with a probability of 1.0 by C4.5, and prior probabilities have no effect in these regions of the feature space. This is distinct from the behavior of maximum likelihood classification where the estimated probabilities for all cases are affected, and low prior probability can preclude selection of a class, even if it is spectrally distinct (Strahler, 1980).

### 3.2. Training sample characteristics and prior probabilities

Fig. 1c illustrates the influence that training sample class frequency distributions can have on classification predictions from nonparametric algorithms. In particular, when classes within a region of feature space are not separable, the distribution of class predictions tends to reflect the class frequency distribution of the training sample (Foody, McCulloch, & Yates, 1995). This bias occurs because nonparametric classification algorithms such as decision trees are optimized to maximize classification accuracy based on the training data provided. They therefore assume that the training sample captures the variability and characteristics of the underlying population, and implicitly assume that the relative class frequency of the training sample is representative of the population (Breiman et al., 1984, p. 9; Brown & Koplowitz, 1979; Ripley, 1996, p. 220).

Fig. 1c illustrates bias in classification predictions arising from variation in the class frequency distribution of the training data. To generate this plot, a series of training data sets were created where a fixed number of samples were randomly drawn from Class 1 while varying the number of samples drawn from Class 2. Fig. 1c shows that the proportion of the test data predicted to belong to Class 2 increases as the proportion of samples belonging to Class 2 in the training data increases. By oversampling Class 2 relative to Class 1, the Class 2 sample density curve is effectively shifted up (i.e., higher probability) relative to the population probability density for Class 2 shown in Fig. 1a. As a result, the point where the two class sample density curves intersect shifts to the left (to a lower value of $X$), and a greater proportion of the feature space is predicted to belong to Class 2. Therefore, oversampling of one or more classes in training data can lead to bias in classification predictions. In Section 5 we show that using prior probabilities provides a means of correcting this type of bias.

### 4. Evaluation using Landsat TM imagery

#### 4.1. Data

The method described in Section 2 was assessed using field data and Landsat TM imagery. To do this, data from TM bands 1–5 and 7 were extracted from imagery acquired on September 8, 1995 for 1013 field-validated sites in the Sierra National Forest in California. These sites were compiled by the United States Forest Service (USFS) as part of a remote-sensing-based classification of the Sierra National Forest (Carpenter et al., 1999; Woodcock et al., 1994). The site data were comprised of 59,903 pixels, each with two class labels. The first label describes the vegetation life form. The second label provides more detailed species information using the CALVEG system (Matyas & Parker, 1980), which classifies vegetation according to dominant species associations. The class labels and number of sites for each class are presented in Table 1. For this work, the nonconifer CALVEG classes were aggregated into a single “nonconifer” class (seven classes total). This was done to reduce confusion caused by undersampled classes (noted by Carpenter et al., 1999), which complicates interpretation of the results.

The USFS classification was performed using a multi-stage process following the methodology described by Woodcock et al. (1994). This methodology used image segmentation to delineate vegetation stands (Shandley, Franklin, & White, 1996; Woodcock & Harward, 1992), followed by supervised classification of remotely sensed imagery to predict the life form for each stand. Terrain-based species association rules were then used to predict CALVEG labels, followed by manual interpretation and relabeling of errors.

#### 4.2. Estimation of prior probabilities

Bayes’ rule requires that prior probabilities be available for all classes of interest. Thus, ancillary information that is available for only a subset of classes is problematic.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of the field-validated training sites for the Sierra National Forest data set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life form</td>
<td>Number of sites</td>
</tr>
<tr>
<td>Conifer</td>
<td>561</td>
</tr>
<tr>
<td>Red fir</td>
<td>116</td>
</tr>
<tr>
<td>Subalpine conifer</td>
<td>37</td>
</tr>
<tr>
<td>Mixed conifer: fir</td>
<td>121</td>
</tr>
<tr>
<td>Eastside pine</td>
<td>22</td>
</tr>
<tr>
<td>Lodgepole pine</td>
<td>41</td>
</tr>
<tr>
<td>Hardwood</td>
<td>211</td>
</tr>
<tr>
<td>Canyon live oak</td>
<td>60</td>
</tr>
<tr>
<td>Blue oak</td>
<td>69</td>
</tr>
<tr>
<td>Buckeye interior live oak</td>
<td>33</td>
</tr>
<tr>
<td>Scrub</td>
<td>38</td>
</tr>
<tr>
<td>Northern mixed chaparral</td>
<td>2</td>
</tr>
<tr>
<td>Mid-elevation chaparral</td>
<td>16</td>
</tr>
<tr>
<td>High-elevation chaparral</td>
<td>3</td>
</tr>
<tr>
<td>Herbaceous</td>
<td>101</td>
</tr>
<tr>
<td>Wet meadow</td>
<td>50</td>
</tr>
<tr>
<td>Water</td>
<td>50</td>
</tr>
<tr>
<td>Barren</td>
<td>50</td>
</tr>
</tbody>
</table>
Following the methodology suggested by Chen et al. (1999), the approach taken here was to assign equal prior probability of \( \frac{1}{J} \) to each of the classes for which no information was available, where \( J \) was the total number of classes. The prior probabilities for the other classes were then taken to be proportional to their likelihood estimates (derived from ancillary information). The final prior probabilities were then normalized to sum to 1.0 for each pixel.

Prior probabilities describing associations between topographic variables (slope, aspect, elevation) and CALVEG labels for conifer sites were estimated from the field data. To estimate these probabilities, a boosted classification tree was constructed using slope, aspect, and elevation data to predict conifer species. This method is similar to the approach followed by Franklin (1998) who demonstrated that useful species association rules can be estimated from terrain data using decision trees. The estimated terrain rules discriminated between species within a single life form (conifers), but not between different life forms, and provided per-pixel prior probabilities of class membership for each conifer species at each pixel in the study area. Conifers were selected because they represent the dominant tree species in the region, comprising 55% of the training sites and 6 of the 18 CALVEG classes.

Two steps were required to estimate the terrain rules. First, prior probabilities for conifer species were produced for the nonconifer training sites. This step was necessary because prior probabilities were required for all classes at each pixel. Note that because the training data used to estimate the boosted decision trees did not include nonconifer sites, the estimated prior probabilities for these sites possessed lower predictive accuracy relative to the conifer sites. Second, estimates of prior probabilities were produced for the conifer training sites using a jackknife cross-validation procedure (Mosteller & Tukey, 1977). In this procedure, boosted decision trees were estimated after removing training sites from the training data, one at a time. In doing so, independent estimates of conifer class membership were obtained for every training site, using only terrain data. To ensure that no classes were eliminated by the terrain rules, the prior probabilities were constrained to be greater than 0 for every class. Following the procedure described above, the prior probability for the nonconifer class was assigned an initial value of \( \frac{1}{J} \), whereupon the prior probabilities for all classes were normalized to sum to 1.0.

4.3. Classification results

Before including prior probabilities, a decision-tree classification was performed using the Landsat TM data. This classification produced a cross-validated accuracy of 59.7%. The cross-validated accuracy for the conifer sites from the terrain rules (i.e., excluding the nonconifer sites and independent of the TM data) was 57.5%. However, the accuracy across the full data set was only 31.7%. These results indicate that the terrain rules provided useful information for classifying conifer species, but much lower predictive power for the nonconifer class.

To incorporate prior probabilities in the TM-based classification a value for the confidence measure (\( c \)) was required. Fig. 2 presents cross-validated accuracies for the TM-based classification using prior probabilities, where the value of \( c \) varies from 0 to 1. This figure shows that when \( c \) is 0, the cross-validated accuracy is the same as that achieved using only spectral data (shown by the horizontal dashed line in Fig. 2). Conversely, because the terrain rules were relatively poor at classifying the nonconifer sites, setting \( c \) equal to 1 produced lower classification accuracy (56.9%) relative to using only spectral data.

The results reported in Fig. 2 illustrate a potential pitfall of using prior probabilities: ancillary information does not necessarily improve classification results. At the same time, because the terrain rules contain useful information for the conifer sites, they can improve classification accuracy if their influence is appropriately weighted. Specifically, Fig. 2 shows that the cross-validated classification accuracy increased to 65.1% when \( c = 0.4 \). This result represents the optimal case where the benefits of including information from the terrain rules are maximized and the costs are minimized. As a consequence, the resulting accuracy is higher than the accuracy obtained using either source of information alone.

Closer inspection of the classification results illustrates these patterns more clearly. Table 2 shows that the accuracies for the conifer classes increase until \( c \approx 0.4 \). As \( c \) increases beyond this threshold, accuracies for the conifer species generally continue to rise. However, because the terrain rules are only effective for the conifer classes, increasing \( c \) (and therefore the overall likelihood of the conifer species), increases the frequency with which nonconifer examples are misclassified as conifers. Thus,
selecting an appropriate value for $c$ helps to balance the useful information in the terrain rules against the errors that they introduce.

Determining an appropriate value for the confidence parameter $c$ is potentially problematic and subjective. However, Fig. 2 shows that the value of $c$ for which the cross-validated accuracy is maximum roughly corresponds to the cross-validated accuracy of the terrain rules for the full training set ($\approx 32\%$, shown by the vertical dashed line in Fig. 2). This suggests that the accuracy of the terrain rules provides a good estimate of their quality, and by extension, a good estimate of $c$. Intuitively, using the accuracy of the terrain rules (or more precisely, the proportion of correctly classified cases) as a value for $c$ makes sense. If the predictive accuracy of the terrain rules is high, their influence should also be high, and vice versa. While the classification results shown in Table 2 are influenced by the choice of $c$, they do not appear to be highly sensitive to the exact value selected. As long as the confidence parameter generally captures the quality of the ancillary information, improvement is achieved.

5. Application to coarse-resolution mapping

Confusion between natural vegetation and agriculture is a major source of error in remote-sensing-based global land cover maps. For example, Loveland et al. (1999) reported that nearly 60% of the problems addressed in the postclassification process for the International Geosphere Biosphere Programme’s Data and Information System (IGBP-DIS) global land cover data set arose from confusion between natural vegetation and agriculture. This problem was also noted in the MODIS prototype classification of North American land cover based on AVHRR data (Friedl et al., 2000). The most obvious confusion in this regard arose because of seasonal variation in the AVHRR NDVI signal for evergreen needle leaf land cover at high latitudes (caused by seasonal variation in illumination geometry) that mimicked a phenological cycle (Spanner, Pierce, Running, & Peterson, 1990). Because of this, the classification algorithm confused this signal with agricultural (and other) land cover. This factor, in conjunction with the limited information content of the AVHRR NDVI data (Goward, Markham, Dye, Dulaney, & Yang, 1991) and oversampling of agriculture in the training data, resulted in overprediction of agriculture in the final map.

One means to address this problem is to use available knowledge concerning the spatial distribution of agriculture to constrain predictions from spectral information alone. Prior probabilities provide a means of incorporating such knowledge. Recently, Ramankutty and Foley (1998) produced global maps of agricultural intensity at 5-min spatial resolution. Their estimates were derived from an analysis of global season land cover regions (Loveland et al., 1999) along with national and regional crop inventory data. While these maps are not perfect, they do provide significant information regarding the global distribution and intensity of agriculture.

For this analysis, agricultural intensities from Ramankutty and Foley (1998) were used to provide likelihood estimates (prior probabilities) for the presence of agriculture in each 1-km pixel. These estimates ranged from 0.01 to 0.95 (the original intensity estimates ranged from 0 to 1.0 but were trimmed to avoid precluding or overpredicting agriculture) and were used in conjunction with the North American land cover training site data set used by Friedl et al. (2000). This data set included 455 sites that were produced by manual interpretation of Landsat TM scenes using the IGBP land cover classification system, which is composed of 17 broad cover types (including two agricultural classes) (Loveland & Belward, 1997).

To incorporate prior probabilities, two steps were required. First, the training site data were used to estimate a boosted decision tree that produced (for each pixel) probabilities of class membership for each IGBP land cover class. Second, Bayes’ rule was used where prior probabilities of agricultural intensity were included as described above. For North America, the agricultural intensity map had a correlation of 0.93 with inventory data from the United States Department of Agriculture (Ramankutty & Foley, 1998). For this work, a somewhat more conservative confidence estimate was used ($c=0.5$).

Table 2
Class accuracies as a function of the confidence parameter $c$

<table>
<thead>
<tr>
<th>Class</th>
<th>0.0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>0.927</td>
<td>0.927</td>
<td>0.912</td>
<td>0.893</td>
<td>0.886</td>
<td>0.814</td>
<td>0.779</td>
<td>0.720</td>
<td>0.652</td>
<td>0.620</td>
<td>0.582</td>
</tr>
<tr>
<td>Lodgepole pine</td>
<td>0.024</td>
<td>0.048</td>
<td>0.097</td>
<td>0.121</td>
<td>0.170</td>
<td>0.170</td>
<td>0.170</td>
<td>0.195</td>
<td>0.170</td>
<td>0.170</td>
<td>0.170</td>
</tr>
<tr>
<td>Eastside pine</td>
<td>0.090</td>
<td>0.136</td>
<td>0.181</td>
<td>0.181</td>
<td>0.318</td>
<td>0.318</td>
<td>0.318</td>
<td>0.318</td>
<td>0.318</td>
<td>0.318</td>
<td>0.363</td>
</tr>
<tr>
<td>Ponderosa pine</td>
<td>0.058</td>
<td>0.088</td>
<td>0.127</td>
<td>0.235</td>
<td>0.323</td>
<td>0.392</td>
<td>0.450</td>
<td>0.519</td>
<td>0.529</td>
<td>0.549</td>
<td>0.578</td>
</tr>
<tr>
<td>Red fir</td>
<td>0.681</td>
<td>0.715</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td>0.724</td>
<td>0.706</td>
<td>0.698</td>
<td>0.689</td>
<td>0.681</td>
<td>0.681</td>
</tr>
<tr>
<td>Subalpine conifer</td>
<td>0.162</td>
<td>0.216</td>
<td>0.378</td>
<td>0.459</td>
<td>0.513</td>
<td>0.594</td>
<td>0.648</td>
<td>0.648</td>
<td>0.675</td>
<td>0.675</td>
<td>0.675</td>
</tr>
<tr>
<td>Mixed conifer: fir</td>
<td>0.214</td>
<td>0.322</td>
<td>0.388</td>
<td>0.388</td>
<td>0.396</td>
<td>0.396</td>
<td>0.388</td>
<td>0.404</td>
<td>0.413</td>
<td>0.413</td>
<td>0.438</td>
</tr>
<tr>
<td>Mixed conifer: pine</td>
<td>0.500</td>
<td>0.524</td>
<td>0.524</td>
<td>0.524</td>
<td>0.549</td>
<td>0.581</td>
<td>0.614</td>
<td>0.622</td>
<td>0.614</td>
<td>0.606</td>
<td>0.622</td>
</tr>
<tr>
<td>Overall</td>
<td>0.597</td>
<td>0.621</td>
<td>0.637</td>
<td>0.643</td>
<td>0.643</td>
<td>0.651</td>
<td>0.639</td>
<td>0.633</td>
<td>0.616</td>
<td>0.585</td>
<td>0.571</td>
</tr>
</tbody>
</table>
When applied to the AVHRR-based classification of North America, the inclusion of prior probabilities resulted in 2.6% of land pixels changing from natural vegetation to agriculture and 6.3% changing from agriculture to natural vegetation. These changes correct many previously noted problems in the original classification (Friedl et al., 2000; Muchoney, 2000). The majority of pixels whose labels changed to agriculture from other classes were located in the agricultural regions of the central and southern United States, while the majority of pixels that were relabeled as natural vegetation were located at high latitudes and in arid regions of the western United States and Mexico.

Fig. 3 illustrates specific examples of classification changes that occurred as a result of including prior probabilities for agriculture for three regions of North America centered in Alaska, Northern California, and South Carolina. The left-hand column of Fig. 3 presents the original predictions from the North American data set, while the right-hand column presents the same areas after including prior probabilities for agriculture (with agriculture shown in red). Dramatic changes can be seen in the Alaska images (top row) where a large area in the center of this region had originally been misclassified as agriculture. When prior probabilities are included, the majority of the agriculture in this region changes to a mix of evergreen needle leaf forest, shrubland, and woody savanna (one of two woodland classes in the IGBP system).

The middle row presents a similar comparison for California and western Nevada. In the left-hand panel, considerable confusion between agriculture and natural vegetation is evident throughout the Sierra Nevada Mountains and in the Great Basin of Nevada. Inclusion of prior probabilities resolves much of this confusion. Also, agricultural area increases in the heavily agricultural region of the Central Valley, and decreases in the suburban areas surrounding San Francisco Bay and in the desert areas of the Great Basin.

Finally, in the two lower panels we present a region where the area mapped as agriculture increases as a result of including prior probabilities for agricultural intensity. Specifically, significant portions of the southeastern United States, previously labeled as evergreen needle leaf forest, were reclassified as agriculture. Once again, these changes resolve known errors in the original map in which significant amounts of agriculture were misclassified as natural vegetation in this region.

Fig. 3. Detailed views of changes arising from including prior probabilities for agriculture in an AVHRR-based IGBP classification for North America. The images on the left are the original predictions. The images on the right show the result of including prior probabilities. The three regions are approximately centered on (from the top): Alaska, northern California, and South Carolina.
In summary, the changes introduced by the use of prior probabilities for agriculture appear to have substantially improved the classification of North American land cover. Given the significant problems posed by confusion between agriculture and natural vegetation in coarse resolution land cover mapping, this approach holds considerable promise for improving land cover characterization over large areas.

6. Discussion and conclusions

In this paper, we have presented a method for including prior probabilities in land cover classifications using nonparametric classification methods. The approach draws from recent research regarding the machine learning algorithm known as boosting. Results from this work demonstrate that inclusion of prior probabilities can be effective in supervised classification of remotely sensed data using nonparametric algorithms. Quantitative and qualitative assessments support the potential of this approach at both fine and coarse spatial resolutions. An important aspect of the method is that knowledge regarding the quality of ancillary information can be incorporated into Bayes’ rule.

One limitation of the proposed approach is that specification of prior probabilities is often problematic. Further, even when useful ancillary information is available, it is often difficult to convert into probabilities. Indeed, a common criticism of Bayesian statistical approaches is the subjectivity involved: Any desired result can be obtained if the appropriate prior probabilities are used (Robert, 1994). The use of the confidence parameter \( c \) partially addresses this problem by reducing the influence of the outside information, thereby mitigating the effects of errors and uncertainty, contingent on the user’s confidence in the quality of the ancillary data.

An important advantage of the method described in this paper is that it can be used to reduce bias in predictions from nonparametric classification algorithms introduced by nonrepresentative training samples. As both this and other research has illustrated (e.g., Foody et al., 1995), the creation of training data used for supervised classification always contains subjective elements, and training sample properties often influence classification results in much the same way as prior probabilities. For unsupervised approaches, the manual labeling process can be subjective (Loveland et al., 1999). Therefore, virtually all classifications contain elements that reflect analyst expectations. While the assumptions underlying training samples are often hidden, the subjective elements of using prior probabilities within the classification process are explicit. Further, prior probabilities can be used to offset bias introduced by a nonrepresentative training sample class frequency distribution. For example, in Section 5 the inclusion of prior probabilities helped to correct classification bias towards overpredicting agriculture caused by overrepresentation of this class in the training data.

Finally, the most compelling reason to use prior probabilities in the classification of remotely sensed data is the ability to include outside knowledge from diverse sources in the classification process. For example, considerable knowledge exists concerning the global distribution of vegetation and land cover types as a function of ecological relationships with climate (e.g., Bailey 1998; Box, 1981; Matthews, 1983; Prentice et al., 1992). However, given the extent of human and natural disturbance regimes (Schimel et al., 1997; Townshend, Justice, Gurney, & McManus, 1991), this information alone is inadequate for global land cover mapping purposes. Remote-sensing-based land cover observations have the potential to produce more accurate representations of actual global land cover. However, remotely sensed data often possess limited ability to discriminate between all vegetation and land cover types of interest (Friedl et al., 2000; Myneni et al., 1995; Running, Loveland, Pierce, Nemani, & Hunt, 1995). The use of prior probabilities, as described in this paper, provides a means to exploit the strengths of both approaches while overcoming some of their individual limitations.

Acknowledgments

This work was supported by NASA Grant NAG5-7218 and NASA contract NAS5-31369. The hard work of the land cover team at BU is gratefully acknowledged, along with the staff of the Global Land 1-km AVHRR project at EROS Data Center who kindly provided the 1-km AVHRR data. Curtis Woodcock provided the USFS data set for the Sierra National Forest. The constructive input of two anonymous reviewers is also gratefully acknowledged.

References


Collins, M., Schapire, R. E., & Singer, Y. (2000). Logistic regression,


