An Iterative Framework for the Improvement of Soft Clusters for Classification

submitted to IPCV '09

Rhonda D. Phillips,[1] **Layne T. Watson,**[1,2] **and Naren Ramakrishnan**[1]
Departments of Computer Science[1] and Mathematics[2], MC 0106
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
{rdphllps, ltw, naren}@cs.vt.edu

Randolph H. Wynne
Department of Forestry, MC 0324
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
wynne@vt.edu

Abstract – *This work introduces an iterative soft cluster refinement method that extensively uses soft cluster evaluation to determine which clusters would discriminate between classes in a classification scheme. This iterative refinement is part of the continuous iterative guided spectral class rejection (CIGSCR) classification method for remotely sensed images. Results indicate that CIGSCR produces good classifications of remotely sensed images, and classification accuracies are higher using the iterative refinement in CIGSCR than classifications derived from soft clustering without iterative refinement.*

Keywords: clustering, functional enrichment, statistical pattern recognition, remote sensing

1 Background

Semisupervised classification has received a good deal of attention in the remote sensing community as remote sensing datasets are characterized by a large number of dimensions (hyperspectral imagery) and limited training data. Semisupervised classification algorithms that involve clustering such as the iterative guided spectral class rejection (IGSCR) algorithm ([1],[2], [3]) have the additional benefit of providing a high level of automation compared to strictly supervised classification algorithms. In remote sensing, informational class categories that make up a classification scheme are defined prior to classification and are identified by humans, whereas spectral classes or clusters have mathematical properties (such as mathematically homogeneous spectral waveforms) and are more difficult for humans to identify. For example, suppose a forest/nonforest classification is desired, and forest and nonforest are the informational class categories. Each informational class is composed of multiple spectral classes that can be used in supervised classification, and the individual spectral classes may not be spectrally similar to each other despite all being part of one informational class. Consider the wide range of tree species that could potentially make up a forest informational class in a particular image. An unsupervised technique such as clustering can identify individual classes that are mathematically homogeneous, and has the additional property of guaranteeing that all types of land cover present in a dataset are represented in the spectral classes (clusters). Both tasks are nontrivial for humans to perform when identifying spectral classes for supervised classification. Therefore semisupervised classification algorithms that involve clustering can automatically identify and label spectral classes, providing significant automation over supervised or unsupervised classification alone.

The purpose of this work is to develop an iterative soft cluster refinement framework, analogous to the framework in IGSCR, that is capable of producing soft classifications of remotely sensed images. This framework will potentially affect classification algorithms that have labeled data and involve clustering. Soft clustering retains all information regarding the proximity of data points to clusters, and will directly produce a soft classification and potentially provide better training spectral classes for a supervised decision rule.

2 IGSCR

IGSCR is a classification method that uses clustering to generate a classification model $p(c_i|x)$ where x is a multivariate sample to be classified and c_i, $i = 1, \ldots, C$, is the ith class where there are C classes in the classification scheme. IGSCR uses clustering to estimate $p(k_j|x)$ in the expression

$$p(c_i|x) = \sum_{j=1}^{K} p(c_i, k_j|x) = \sum_{j=1}^{K} p(c_i|k_j, x)p(k_j|x), \quad (1)$$

where k_j, $j = 1, \ldots, K$, is the jth cluster out of K total clusters. IGSCR also uses the clusters to train a decision rule using Bayes' theorem [4]

$$p(k_j|x) = \frac{p(x|k_j)p(k_j)}{\sum_{i=1}^{K} p(x|k_i)p(k_i)}. \quad (2)$$

The prior probabilities of the clusters $p(k_j)$ are assumed to be equal.

Clustering is performed using a discrete clustering method such as k-means that minimizes the objective function

$$J(\rho) = \sum_{i=1}^{n}\sum_{j=1}^{K} w_{ij}\rho_{ij} \quad (3)$$

subject to

$$\sum_{j=1}^{K} w_{ij} = 1$$

where $w_{ij} \in \{0, 1\}$ is the value in the ith row and jth column of the partition matrix $W \in \Re^{n\times K}$, $U^{(j)} \in \Re^B$ is the prototype for the jth cluster k_j, $x^{(i)} \in \Re^B$ is the ith data point, and $\rho_{ij} = ||x^{(i)} - U^{(j)}||_2^2$. The clusters $k_1, \ldots, k_K$ form a partition of $\{x^{(i)}\}_{i=1}^n$. The algorithm for k-means requires K initial cluster prototypes and iteratively assigns each sample to the closest cluster using

$$w_{ij} = \begin{cases} 1, & \text{if } j = \underset{1\le j\le K}{\operatorname{argmin}}\, \rho_{ij}, \\ 0, & \text{otherwise,} \end{cases}$$

followed by the cluster prototype (mean) recalculation

$$U^{(j)} = \sum_{i=1}^{n}(w_{ij}x^{(i)}) \Big/ \sum_{i=1}^{n} w_{ij}$$

once W has been calculated [5]. This process, guaranteed to terminate in a finite number of iterations, continues until no further improvement is possible, terminating at a local minimum point of (3).

IGSCR uses labeled data in a semisupervised clustering framework to locate clusters that correspond to classes in a given classification scheme. IGSCR requires a labeled set of training data comprised of individual samples within the image to be classified and corresponding class labels. Rather than using the labeled data to train a decision rule directly, the entire image is clustered, thereby capturing the inherent structure of all the data and not just the labeled samples. The clusters represent spectral classes, and in remote sensing, each spectral class ideally maps to exactly one class in the final classification scheme. Once clusters are generated, each cluster must be assigned to one class or rejected as impure. While theoretically each cluster should contain samples belonging to only one informational class, in practice clusters (spectral classes) that predominantly contain samples of one class can contain a few samples from other classes because of inherent errors. However, if a cluster contains too many samples from different classes, the cluster itself is considered confused and should not be labeled with one class. Impure clusters are rejected and can be further refined in the iterative part of the algorithm.

The homogeneity test for cluster purity is performed using the labeled training set. IGSCR produces a hard classification and uses a discrete clustering method where each sample is assigned to exactly one cluster. Let $V_{c,j}$ be the binomial random variable denoting the number of labeled samples assigned to the jth cluster that are labeled with a particular cth class. Let p be the user-supplied cluster homogeneity threshold ($p = .9$ would indicate a cluster is 90% pure with respect to the majority class), and let α be the user- supplied acceptable one-sided Type-I error for a statistical hypothesis test. Then if c is the majority class represented in the jth cluster, the jth cluster is rejected if $P(Z < \hat{z}) < 1 - \alpha$ where Z is a standard normal random variable, m is the number of labeled samples in the jth cluster, and

$$\hat{z} = \frac{v_{c,j} - mp}{\sqrt{mp(1-p)}}. \quad (4)$$

(Typically a continuity correction of 0.5 is added in the numerator of (4).)

If a cluster is rejected, the samples making up that cluster can be reclustered in subsequent iterations. All samples belonging to pure clusters are removed from the image being clustered, resulting in only samples belonging to impure clusters being reclustered. Once more clusters are generated, those clusters are evaluated for purity, removed from the image, and clustering is performed again until termination criteria are met. All samples can belong to pure clusters, leaving no remaining samples to be clustered, no pure clusters could be found in the previous iteration, meaning that the clustering would continue to be performed on the same data, resulting in the same impure clusters (assuming deterministic cluster seeding), or a set number of iterations can be reached, resulting in termination of the iteration. Note that deterministic seeding ensures that the iteration will terminate, even without specifying a maximum number of iterations.

Once the iterative clustering is complete, one or more classifications is performed. The first classification is called the iterative stacked (IS) classification because it is the result of combining or "stacking" all cluster assignments over all iterations (each sample will be assigned to at most one accepted cluster). Assume

that all samples not assigned to an accepted cluster are combined to form one cluster k_{K+1}, and the class assignment for that cluster is "unclassified" or c_{C+1}. Then the IS assignment for a pixel using (1) is

$$\begin{aligned}\mathrm{IS}(x) &= \underset{1\le i\le C+1}{\operatorname{argmax}}\, p(c_i|x) \\ &= \underset{1\le i\le C+1}{\operatorname{argmax}} \sum_{j=1}^{K+1} p(c_i|k_j,x)p(k_j|x),\end{aligned}$$

where

$$p(c_i|k_j,x) = \begin{cases} 1, & \text{if } k_j \text{ is labeled } c_i, \\ 0, & \text{otherwise,}\end{cases}$$

and

$$p(k_j|x) = \begin{cases} 1, & \text{if } x \in k_j, \\ 0, & \text{otherwise,}\end{cases}$$

since cluster assignments are discrete.

The second possible classification, the decision rule (DR) classification, uses the pure clusters to form a decision rule. Recall in (2) that

$$p(k_j|x) = \frac{p(x|k_j)}{\sum_{i=1}^{K} p(x|k_i)}$$

when all the $p(k_j)$ are equal. Traditionally, the maximum likelihood decision rule, assuming a multivariate normal distribution

$$p(x|k_j) = 2\pi^{-B/2}|\Sigma_j|^{-1/2}e^{-\frac{1}{2}(x-U^{(j)})^T\Sigma_j^{-1}(x-U^{(j)})},$$

is used where Σ_j is the covariance matrix of the jth cluster [6]. Since IGSCR produces hard classifications, the full probability need not be calculated as determining only the cluster associated with the maximum probability is necessary. The DR classification function is

$$\begin{aligned}\mathrm{DR}(x) &= \underset{1\le i\le C}{\operatorname{argmax}}\, p(c_i|x) \\ &= \underset{1\le i\le C}{\operatorname{argmax}} \sum_{j=1}^{K} p(c_i|k_j,x)p(k_j|x),\end{aligned} \tag{5}$$

where

$$p(k_j|x) = \begin{cases} 1, & \text{if } j = \underset{1\le j\le K}{\operatorname{argmax}} \big(-\ln|\Sigma_j| \\ & \qquad -(x-U^{(j)})^T\Sigma_j^{-1}(x-U^{(j)})\big), \\ 0, & \text{otherwise.}\end{cases}$$

A final classification, the iterative stacked plus (IS+) classification, combines the DR and IS classifications. If a sample is labeled as unclassified in the IS classification, the DR class value is used for the IS+ classification, otherwise the IS class value is used for that particular sample. The IS+ classification function is

$$\mathrm{IS}+(x) = \begin{cases} \mathrm{IS}(x), & \text{if } x \notin k_{K+1}, \\ \mathrm{DR}(x), & \text{otherwise.}\end{cases}$$

3 CIGSCR

Continuous IGSCR (CIGSCR) uses a similar semisupervised clustering framework to the one established in IGSCR to produce a soft or probabilistic classification instead of a hard classification, and uses continuous algorithms and models instead of discrete algorithms and models. Recall in (1) that $p(c_i|k_j,x)$ and $p(k_j|x)$ are either 0 or 1 (discrete) in practice in IGSCR. $p(c_i|k_j,x)$ is necessarily discrete because while several clusters can comprise one class, only one class (theoretically) can label the members of a particular cluster, but there are no similar restrictions on $p(k_j|x)$. In fact, the clustering algorithm and the maximum likelihood decision rule indicate positive probabilities that a sample is associated with each cluster, but IGSCR makes an assignment only to the cluster with the highest probability.

Consider a soft clustering algorithm that minimizes the objective function [7]

$$\begin{aligned}J(\rho) = \sum_{i=1}^{n}\sum_{j=1}^{K} w_{ij}^p \rho_{ij} \quad &\text{subject to} \\ \sum_{j=1}^{K} w_{ij} = 1 \text{ for each } i&\end{aligned} \tag{6}$$

where $w_{ij} \in (0,1)$ is the value in the ith row and jth column of the weight matrix $W \in \Re^{n\times K}$ (analogous to the partition matrix W in (3)), $U^{(j)} \in \Re^B$ is the jth cluster prototype, $p > 1$, and $\rho_{ij} = \rho(x^{(i)}, U^{(j)}) = ||x^{(i)} - U^{(j)}||_2^2$ is the Euclidean distance squared. The algorithm that minimizes this objective function is similar to that of k-means in that it first calculates

$$w_{ij} = \frac{(1/\rho_{ij})^{1/(p-1)}}{\sum_{k=1}^{K}(1/\rho_{ik})^{1/(p-1)}}$$

for all i and j followed by calculating updated cluster prototypes

$$U^{(j)} = \sum_{i=1}^{n} w_{ij}^p x^{(i)} \Big/ \sum_{i=1}^{n} w_{ij}^p.$$

This iteration (recalculation of the weights followed by recalculation of cluster prototypes, following by recalculation of the weights, etc.) is guaranteed to converge (with these definitions of ρ_{ij}, $U^{(j)}$, and w_{ij}) for $p > 1$ [8].

The classification function for IS classification is

$$\mathrm{IS}(x) = p(c_i|x) = \sum_{j=1}^{K} p(c_i|k_j,x)p(k_j|x), \tag{7}$$

where $p(k_j|x)$ is estimated using w_{ij} and $p(c_i|k_j,x)$ does not change from IGSCR. The classification function for the DR classification is

$$\mathrm{DR}(x) = p(c_i|x) = \sum_{j=1}^{K} p(c_i|k_j,x)p(k_j|x)$$

$$= \frac{\sum_{j=1}^{K} p(c_i|k_j,x)\left[\frac{2e^{-\frac{1}{2}(x-U^{(j)})^T\Sigma_j^{-1}(x-U^{(j)})}}{\pi^{B/2}|\Sigma_j|^{1/2}}\right]}{\sum_{l=1}^{K}\left[\frac{2e^{-\frac{1}{2}(x-U^{(l)})^T\Sigma_l^{-1}(x-U^{(l)})}}{\pi^{B/2}|\Sigma_l|^{1/2}}\right]}. \quad (8)$$

An analog for the IS+ classification is unnecessary in CIGSCR as all samples will be part of pure clusters and will be classified.

4 Association Significance Test

Clusters will be evaluated using the association significance test developed in a companion paper. The association significance test is based on a statistical hypothesis test using the Wald statistic

$$\hat{z} = \frac{y_{c,j} - n_c\overline{w}_j}{\sqrt{p_c\sum_{d=1}^{C} n_d\big(S^2_{\overline{w}_{d,j}} + (1-p_c)\overline{w}^2_{d,j}\big)}}, \quad (9)$$

where $y_{c,j}$ is the sum of cluster weights for samples labeled with the cth class to the jth cluster, n_c is the number of samples labeled with the cth class, $\overline{w}_j$ is the sample mean of all weights in the jth cluster, p_c is $\frac{n_c}{n}$, n is the number of labeled samples,

$$\overline{w}_{c,j} = \frac{1}{n_c}\sum_{i\in J_c} w_{ij},$$

$$S^2_{\overline{w}_{c,j}} = \frac{1}{n_c - 1}\sum_{i\in J_c}(w_{ij} - \overline{w}_{c,j})^2,$$

and J_c is the index set of all samples labeled with the cth class.

The null hypothesis that the average cluster weights for the cth class are not statistically significantly different from other cluster weights in the jth cluster is rejected if $P(Z > \hat{z}) < \alpha$, meaning the cluster is not rejected.

5 Iteration

Together with the cluster association significance test, the iteration forms the clustering framework in CIGSCR. The application of a hypothesis test determines which clusters should be used for classification, and an iteration works to produce a set of associated clusters with each class being represented by at least one associated cluster. This is accomplished by introducing new clusters that are likely to be associated, and when necessary, are associated with a class not already represented by a cluster.

In IGSCR, pure hard clusters are removed from the image that is clustered in subsequent iterations, focusing further refinement on clusters that failed to pass the purity test. K clusters are used for each iteration, presumably producing smaller clusters as less data is divided into the same number of clusters. The underlying assumption is that clusters that fail to pass the purity test could actually be composed of multiple clusters that would pass the purity test individually, and clustering the remaining data into K more clusters will reveal these smaller clusters. This method will not directly work on soft clusters as soft clusters cannot be removed simply by removing any sample associated with a pure cluster—all samples have a positive probability of belonging to any particular cluster.

In CIGSCR, unassociated clusters are targeted for refinement by using their information to create new clusters that will likely be associated. IGSCR is effectively locating smaller clusters that when combined to form a larger cluster would have been rejected. IGSCR accomplishes this by finding the same number of clusters (K) in the original dataset and then in successively smaller subsets of that original dataset. A similar approach that would locate smaller pure clusters in rejected clusters is "splitting" a cluster, employed by Ball and Hall [9] in ISODATA. Clusters are split by partitioning a cluster into two new clusters and recalculating new means. Soft clusters are represented by cluster means, and splitting a soft cluster would equate with replacing one cluster mean with two cluster means (calculated based on data associated with a cluster).

A cleaner algorithmic solution is to add one new cluster using information contained in the target cluster (the cluster that would be split), which effectively splits the cluster into two clusters. When using a clustering algorithm based on objective function (6), adding a new cluster guarantees a smaller function value (shown below) when $p = 2$. Using only the labeled samples belonging to the majority class (as determined in the cluster association significance test) to seed a new cluster would have the effect of pulling the new cluster toward those samples. Once another clustering iteration is completed, the targeted cluster would produce one cluster that is likely to be associated with the majority class and another cluster that retains relatively strong associations with all other classes. In CIGSCR, once the association significance test is performed, if at least one cluster is unassociated (and there are no unassociated classes), the cluster with the lowest value of $\hat{z}$ is used

to generate a new cluster. The new cluster mean is determined using

$$U^{(K+1)} = \frac{\sum_{i \in J_{c_k}} w_{ik} X^{(i)}}{\sum_{i \in J_{c_k}} w_{ik}}, \tag{10}$$

where k is the cluster with the lowest value of $\hat{z}$, c_k is the majority class in cluster k, and recall that J_c is the index set of labeled samples whose label is c. This formula also works when a class other than the majority class is used to seed a new cluster mean.

A shortcoming in IGSCR is that there is no guarantee that any clusters will be created and labeled with any particular class, and if a particular class is not represented by a cluster, the desired classification cannot be performed. In CIGSCR, this issue is addressed by adding a new cluster using information from a particular class if that class is not represented in the associated clusters. If a class c is not represented in the associated clusters, the cluster that is closest to being associated with c is used to generate a new cluster using (10) with $c_k = c$. The "closest" cluster is determined to be the cluster with the highest ratio of the average membership of class c to the average membership of the majority class.

When there are classes not represented by associated clusters and there are unassociated clusters, only one method can be used to determine the creation of a new cluster. If a cluster is unassociated, it is simply not used in classification. It is more important to have each class represented by the associated clusters than to refine an unassociated cluster, because the desired classification cannot be applied unless all classes are represented by associated clusters. Therefore adding a new cluster so that all classes will be represented takes precedence over adding a new cluster because an existing cluster is unassociated.

Finally, the theorem proving that adding one cluster mean will result in a smaller value of (6) is presented below.

Theorem: Given an integer $K > 0$, positive real numbers ρ_{ij}, $i = 1, \ldots, n$; $j = 1, \ldots, K+1$, defining a point $\rho \in \Re^{n \times K+1}$, and the objective function

$$J^{(K)}(\rho) = \sum_{i=1}^{n} \sum_{j=1}^{K} w_{ij}^2 \rho_{ij},$$

for K clusters where

$$w_{ij} = \frac{1/\rho_{ij}}{\sum_{k=1}^{K} 1/\rho_{ik}},$$

the objective function

$$J^{(K+1)}(\rho) = \sum_{i=1}^{n} \sum_{j=1}^{K+1} \hat{w}_{ij}^2 \rho_{ij},$$

for $K+1$ clusters where

$$\hat{w}_{ij} = \frac{1/\rho_{ij}}{\sum_{k=1}^{K+1} 1/\rho_{ik}},$$

satisfies

$$J^{(K+1)}(\rho) < J^{(K)}(\rho).$$

Proof: Note that the ρ_{ij} do not change with the addition of the $(K+1)$st cluster prototype, however $\hat{w}_{ij} < w_{ij}$ for $j < K+1$ because the denominator of $\hat{w}_{ij}$ has an additional term. Let $J_i^{(K)} = \sum_{j=1}^{K} w_{ij}^2 \rho_{ij}$ and $J_i^{(K+1)} = \sum_{j=1}^{K+1} \hat{w}_{ij}^2 \rho_{ij}$. It is sufficient to show that $J_i^{(K+1)} < J_i^{(K)}$ for each i to prove that $J^{(K+1)} < J^{(K)}$.

Let

$$S_1 = \sum_{k=1}^{K} 1/\rho_{ik} \qquad \text{and} \qquad S_2 = \sum_{k=1}^{K+1} 1/\rho_{ik}.$$

Then

$$w_{ij}^2 = \frac{(1/\rho_{ij})^2}{S_1^2} \qquad \text{and} \qquad \hat{w}_{ij}^2 = \frac{(1/\rho_{ij})^2}{S_2^2}.$$

$$\begin{aligned} J_i^{(K)} - J_i^{(K+1)} &= \sum_{j=1}^{K} \frac{(1/\rho_{ij})}{S_1^2} - \sum_{j=1}^{K+1} \frac{(1/\rho_{ij})}{S_2^2} \\ &= \frac{S_2^2 \sum_{j=1}^{K} (1/\rho_{ij}) - S_1^2 \sum_{j=1}^{K+1} (1/\rho_{ij})}{S_1^2 S_2^2}. \end{aligned}$$

Examining only the numerator in the previous term,

$$\begin{aligned} &(S_1 + (1/\rho_{i,K+1}))^2 \sum_{j=1}^{K} (1/\rho_{ij}) \\ &\qquad - S_1^2 \left(\sum_{j=1}^{K} (1/\rho_{ij}) + (1/\rho_{i,K+1}) \right) \\ &= (S_1 + (1/\rho_{i,K+1}))^2 S_1 - S_1^2 (S_1 + (1/\rho_{i,K+1})) \\ &= S_1^3 + 2S_1^2 (1/\rho_{i,K+1}) + S_1 (1/\rho_{i,K+1})^2 \\ &\qquad - S_1^3 - S_1^2 (1/\rho_{i,K+1}) \\ &= S_1^2 (1/\rho_{i,K+1}) + S_1 (1/\rho_{i,K+1})^2 \\ &> 0 \end{aligned}$$

yielding

$$J_i^{(K+1)} < J_i^{(K)}.$$

Q.E.D.

Assuming that the clustering algorithm locates a local minimum point of the objective function, the combination of the clustering algorithm and this cluster prototype addition are guaranteed to move toward a smaller objective function value. If left unchecked, infinitely many clusters could be added, and the algorithm would continue to find smaller objective function values. The association significance test plays a crucial role in the termination of this iterative process. Once all clusters pass the association significance test and each class has at least one associated cluster, the iteration stops because the higher level objective has been met: clusters that significantly correspond to all classes have been located. The iteration also terminates when a maximum number of clusters is reached, and only those clusters that pass the association significance test are used for classification.

6 Experimental Results and Discussion

The dataset used to obtain experimental results for IGSCR and CIGSCR is a mosaicked Landsat Thematic Mapper (TM) satellite image taken from Landsat Worldwide Reference System (WRS) path 17, row 34, located in Virginia, USA, shown in Figure 1. This image, hereafter referred to as VA1734, was acquired on November 2, 2003 and consists largely of forested, mountainous regions, and a few developed regions that are predominantly light blue and light pink in Figure 1. Figure 1 contains a three color representation of VA1734 where the red color band in Figure 1 corresponds to the near infrared wavelength in VA1734, the green color band in Figure 1 corresponds to the red wavelength in VA1734, and the blue color band in Figure 1 corresponds to the green wavelength in VA1734.

The training data for this image was created by the interpretation of point locations from a systematic, hexagonal grid over Virginia Base Mapping Program (VBMP) true color digital orthophotographs. A two class classification was performed (forest/ nonforest), and classification parameters and results are given in Tables 1 and 2. Classification images for this dataset are given in Figures 2 (DR image) and 3 (IS image).

Validation data in the form of point locations at the center of USDA Forest Service Forest Inventory and Analysis (FIA) ground plots were used to assess the accuracy of this classification. Since these validation data are typically used to evaluate crisp classifications, only homogeneous FIA plots were used (either 100 percent forest or nonforest), and these plots were obtained between 1997 and 2001. Accuracy was assessed based on an error matrix where classification results for specific points (not included in the training data set) are compared against known class values. The accuracies

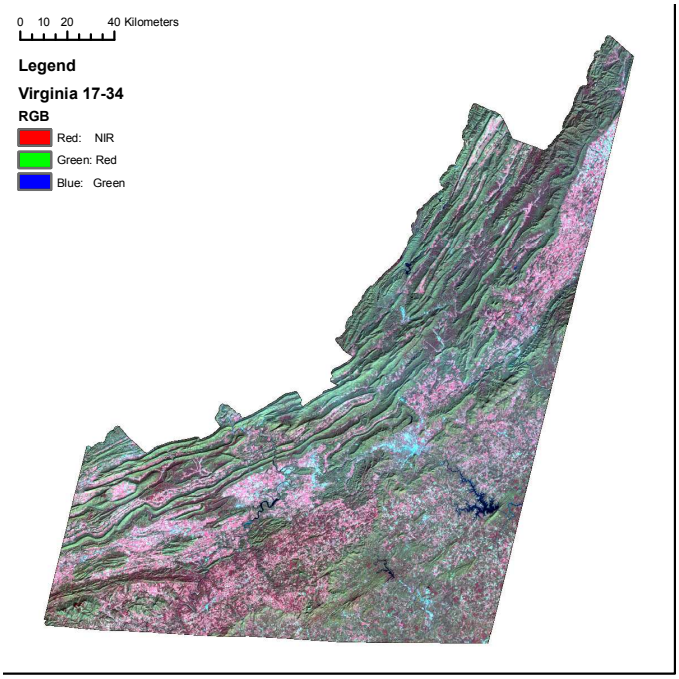

Figure 1. Landsat ETM+ path 17/row 34 over Virginia, USA with area of interest highlighted..

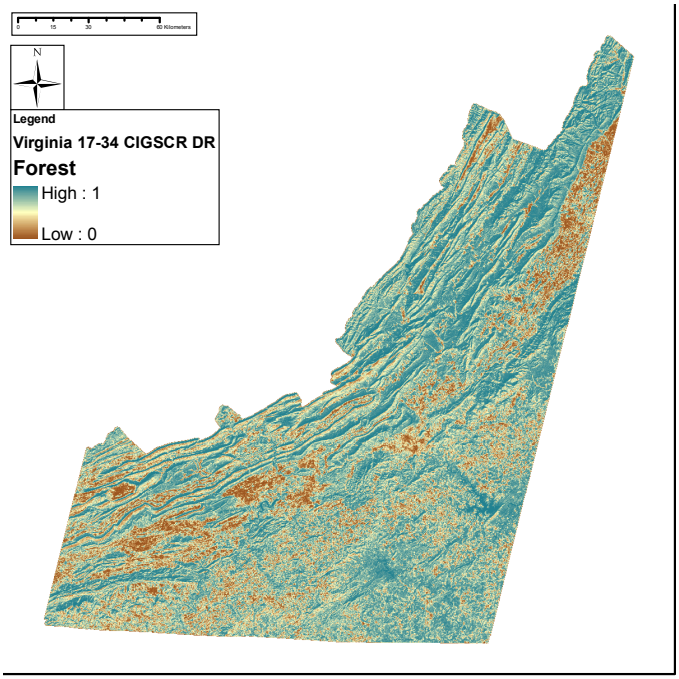

Figure 2. CIGSCR DR classification using 10 initial clusters..

reported in Tables 1 and 2 were obtained by first converting all soft classifications to hard classifications for the purpose of comparing hard classification values to hard ground truth values.

The soft clustering and soft classification in CIGSCR can result in qualitatively different classifications than

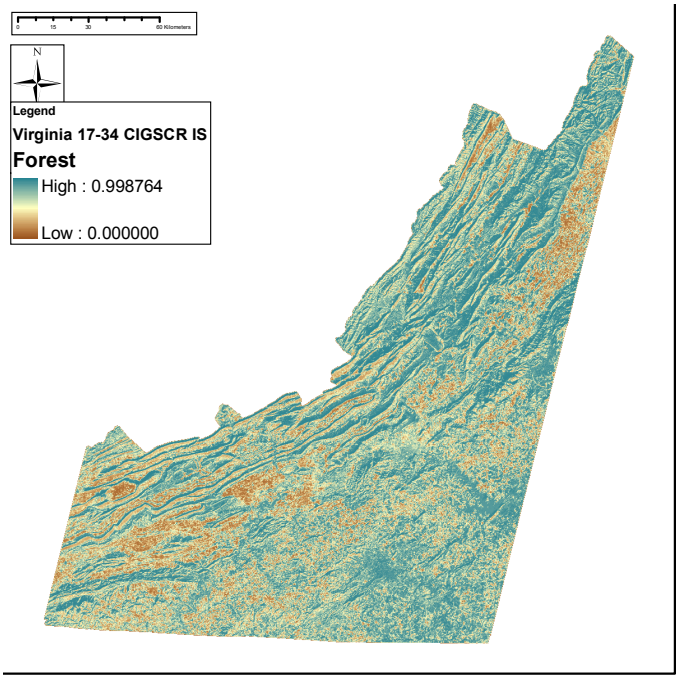

Figure 3. CIGSCR IS classification using 10 initial clusters..

Table 1. IGSCR and CIGSCR decision rule (DR) classification accuracies for VA1734.

no. init.	IGSCR	CIGSCR	clustering
clusters	$p = .9, \alpha = .01$	$\alpha = .0001$	(no iteration)
10	75.49	88.74	72.26
15	74.56	80.50	73.72
20	89.57	79.87	76.54
25	84.25	81.44	77.58

Table 2. IGSCR iterative stacked plus (IS+) and CIGSCR iterative stacked (IS) classification accuracies for VA1734.

no. init.	IGSCR	CIGSCR	clustering
clusters	$p = .9, \alpha = .01$	$\alpha = .0001$	(no iteration)
10	75.39	83.63	72.26
15	74.56	76.96	72.99
20	88.95	75.60	76.85
25	83.94	78.52	76.75

IGSCR. Even when the final classifications are similar, CIGSCR provides more information through soft classification. Furthermore, based on accuracies reported in Tables 1 and 2, CIGSCR is less sensitive to the number of initial clusters than IGSCR. As shown in Tables 1 and 2, IGSCR can be sensitive to the number of initial clusters. The set of clusters ultimately used for classification in IGSCR is directly affected by the number of initial clusters and the homogeneity test, and furthermore, when all clusters fail the homogeneity test, the iteration terminates and no more clusters are found. The number of clusters used for classification can vary widely depending on the number of iterations completed as each iteration potentially produces several pure clusters. The classification accuracies reported for CIGSCR in Tables 1 and 2 are more consistent as CIGSCR does not have the same sensitivity issues. First, the association significance test no longer requires a user input threshold like the homogeneity test. The homogeneity test evaluates the observed values against a user supplied probability of observing a specific class (within a cluster), but the association significance test determines if the average cluster memberships per class are statistically significantly different (requiring no user specified probability). Secondly, the iteration in CIGSCR is fundamentally different from the iteration in IGSCR. While each iteration in IGSCR locates multiple clusters, each iteration in CIGSCR adds one additional cluster, and terminating this iteration potentially excludes many fewer clusters from the final classification than terminating the iteration in IGSCR (especially when few iterations occur). As classification methods are already sensitive to training data and clustering methods are sensitive to initial prototype locations, classifications being sensitive to fewer parameters is a desirable property.

Perhaps the most important question about this iterative clustering scheme is whether using the combination of the association significance test and the iteration improves the clustering for the purposes of classification. Each cluster is labeled with the class that has the highest average membership in the cluster. Observe in experimental runs in Tables 1 and 2 that **all** classification accuracies using just clustering are lower than corresponding classification accuracies using CIGSCR. Based on the available results in Tables 1 and 2, the semisupervised clustering scheme in CIGSCR improves classification accuracies when training data are available to influence clustering.

7 Conclusions

This paper introduces an iterative soft clustering mechanism for soft classification. By incorporating a soft cluster evaluation method, the association significance test presented in a companion paper, into the iteration proposed in this paper, clusters are iteratively improved for the purposes of classification. Results in this paper demonstrated that a soft CIGSCR classification method based on the association significance test and iteration could produce classifications that contain more information than IGSCR classifications. CIGSCR classifications are less sensitive to the number of initial clusters than IGSCR classifications. Finally,

classification accuracies were higher using the iterative clustering method proposed in this paper than just soft clustering with no refinement. The iterative clustering refinement mechanism proposed in this paper has the potential to affect other classification methods that utilize clustering.

References

[1] J.P. Wayman, R.H. Wynne, J.A. Scrivani, and G.A. Reams, "Landsat TM-based forest area estimation using Iterative Guided Spectral Class Rejection," *Photogrammetric Engineering & Remote Sensing*, 67(2001), 1155-1166.

[2] R.F. Musy, R.H. Wynne, C.E. Blinn, J.A. Scrivani, and R.E. McRoberts, "Automated Forest Area Estimation via Iterative Guided Spectral Class Rejection," *Photogrammetric Engineering & Remote Sensing*, 72(2006), 949-960.

[3] R.D. Phillips, L.T. Watson, and R.H. Wynne, "Hybrid image classification and parameter selection using a shared memory parallel algorithm," *Computers & Geosciences*, 33(2007), 875-897.

[4] B.V. Gnedenko, *Theory of Probability (sixth ed.)*, Gordan and Breach Science Publishers, The Netherlands, 1997, 497pp.

[5] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, SIAM, Philadelphia, 2007, 466 pp.

[6] J.A. Richards and X. Jia, *Remote Sensing Digital Image Analysis (third ed.)*, Springer-Verlag, Berlin, 1999, 363 pp.

[7] J. Bezdek, "Fuzzy mathematics in pattern classification," PhD Thesis, Cornell University, Ithaca, NY, 1974.

[8] J.C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(1), 1–8.

[9] Ball, G.H. and Hall, D.J., 1965, "A novel method of data analysis and pattern classification," Stanford Research Institute, Menlo Park, CA.