AN INDUCTION-BASED MODEL FOR CLASSIFICATION OF LANDSAT DATA

Richard J. Roiger, PhD
IN%"ROIGER@VAX1.MANKATO.MSUS.EDU"

Lee D. Cornell, MS
IN%"CORNELL@VAX1.MANKATO.MSUS.EDU"

Associate Professors
Department of Computer and Information Sciences
Mankato State University
Mankato, MN 56002-8400 USA
ISPRS Commission III

The current research presents an induction-based empirical model that uses a heuristic evaluation function capable of utilizing the most predictive attributes in performing classification of satellite data. This paper discusses the structure of this model and compares its classification accuracy and other characteristics to those exhibited by other systems, both heuristic and statistical.

The model is used to analyze Landsat data and perform classification of pixels into one of fifteen different categories, with a demonstrated accuracy rate approaching 100 percent.

Key Words: Artificial Intelligence, Classification, Image Analysis, Landsat

ACKNOWLEDGEMENT

We would like to acknowledge the assistance of Daniel Civco (Civco 1991, 1992a, 1992b) of the University of Connecticut in sharing the test data which he used in the training and testing of a neural net system designed for remote sensing data analysis and classification. These data are sampled image data derived from a May 1988 Landsat Thematic Mapper (TM) scene consisting of multispectral reflectance values in six bands of the electromagnetic spectrum (blue, green, red, near infrared, and two middle infrared) for 15 different land covers.

The availability of these data provided us with the ability to have valid benchmarks in terms of the classification accuracy of the system under development, without having to attempt to repeat work which is already underway by others.

We would also like to acknowledge the continuing support of Dr. Maria Gini, Department of Computer and Information Sciences, at the University of Minnesota.

1. INTRODUCTION

In this paper we present SX-WEB, an exemplar-based concept learning model capable of analyzing digitized satellite images of the earth's surface. SX-WEB is a modification of EX-WEB (Roiger, 1991), an incremental concept formation model of concept learning. With EX-WEB, learning is unsupervised and incremental. An unsupervised paradigm is, in general, inappropriate for image classification since most data images will not contain a representative sampling of all available classification categories. Because of this, learning with SX-WEB is supervised. SX-WEB retains EX-WEB's ability to learn incrementally and to limit the use of the attributes used for classification to those deemed most predictive of class membership. However, for rapid classifications, SX-WEB is best used as a non-incremental system. SX-WEB can classify in domains containing nominal, real-valued and mixed data (both nominal and real-valued data exist). Because digitized images are real-valued, we will concentrate on SX-WEB's real-valued data structure and similarity measure.

SX-WEB is written in PC Scheme. Scheme is a LISP-based language conceived in the 1970s at MIT by G.L. Steele and G.J. Sussman. PC Scheme is an adaptation of Scheme developed by Texas Instruments in the 1980s.

The training and testing data which was provided by Daniel Civco consisted of 302 pixels for which ground truth had been established. These data had been classified into fifteen categories: Urban (UR), Agriculture 1 (A1), Agriculture 2 (A2), Turf/Grass (TG), Southern Deciduous (SD), Northern Deciduous (ND), Coniferous (CO), Shallow Water (SW), Deep Water (DW), Marsh (MA), Shrub Swamp (SS), Wooded Swamp (WS), Dark Barren (DB), Barren 1 (B1), and Barren 2 (B2). Each pixel was represented by six values, consisting of the multispectral reflectance values in six bands of the electromagnetic spectrum: blue (0.45-0.52 μm), green (0.52-0.60 μm), red (0.63-.069 μm), near infrared (0.76-0.90 μm), and two middle infrared (1.55-1.75 and 2.08-2.35 μm).

2. THE SX-WEB LEARNING MODEL

In this section, we examine in detail the main features of SX-WEB with help from the domain of Landsat data images. We present SX-WEB's exemplar-based similarity measure and evaluation function. We conclude this section with a complexity analysis.

2.1 Representing real-valued data with SX-WEB

The primary data structure used by SX-WEB is a three level tree. Figure 1 shows the general form of this tree structure. The nodes at the instance-level of the tree represent the individual training instances that have been used to define the concept classes given at the concept-level. For the domain in question, each instance-level node contains an attribute-value list consisting of the spectral band identifications together with their specific values. The values found within the attributes of the instance nodes are used by SX-WEB's exemplar-based evaluation function to classify newly presented instances whose classification is unknown.

The concept-level nodes of the tree in Figure 1 store the means and standard deviations of the attributes found within their respective instance-level children. That is, concept C_1 contains the means and standard deviations for the attributes found within I_1, I_2, I_3 and I_4. Figure 2 shows the mean and standard deviation scores for the root-level node and the fifteen concept-level classes formed with a training set containing 155 instances. SX-WEB uses these mean and standard deviation scores to determine those attributes most predictive of class membership.

To illustrate this, consider Figure 2 and the mean values for the attribute BLUE. The smallest mean score for the attribute BLUE is 71.4 and is found in the concept class

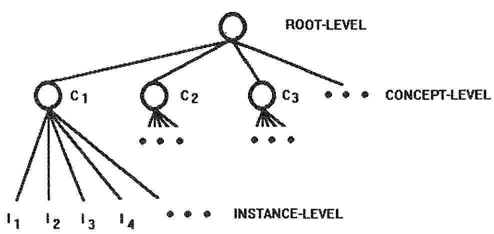

Figure 1: The three-level tree structure of SX-WEB.

representing SHALLOW WATER (SW). The largest mean value for BLUE is 149 and is found in the concept class BARREN 2 (B2). To determine whether BLUE is an attribute predictive of class membership, the standard deviation of the attribute BLUE in the root node (sd=20.63) is used to normalize the mean scores of all concept-level children. That is, each pair of mean values for the chosen attribute are subtracted from one another. The absolute value of each subtraction is then divided by the standard deviation of the attribute found within the root node. These standardized difference values are summed and divided by the number of paired attribute computations that have been made. This gives an average standardized mean difference value for the attribute relative to the root node. If this average difference is larger than a user specified threshold value, the attribute is considered to be predictive. Making this computation for the attribute BLUE results in a value of approximately 1.075. If the predictiveness threshold is set at 1.0, BLUE is then determined to be a predictive attribute. Figure 3 shows the predictiveness scores for the attributes found within the 155 instance training set.

Finally, family resemblance scores are stored within the root node and each concept-level node. Family resemblance scores computed from the 155 instance training set are shown in Figure 2 in the final column, labeled FR. Family resemblance scores form the basis of SX-WEB's evaluation function by giving a measure of the overall similarity of the exemplars making up individual concept classes. The concept class with the lowest family resemblance score is Shrub Swamp (SS). As we will see in Section 2.3, those concept classes containing highly similar instances will have lower family resemblance scores.

2.2 Computing the similarity of two exemplars

SX-WEB uses two formulas to compute similarity. One formula is used when attributes are nominal or mixed and a second similarity measure is used when attributes are strictly real-valued. Once again, we will limit our discussion to real-valued exemplar similarity.

To compute the similarity between exemplars E_1 and E_2, the absolute value of the difference between each attribute value in E_1 and its corresponding attribute value in E_2 is divided by the standard deviation of the attribute found in the concept-level node being considered for instance classification. These standardized differences are summed over all attributes. Finally, the sum of the standardized differences are divided by the number of attributes giving an average standardized difference value among the attributes of E_1 and E_2. Notice that similarity scores closer to zero mean greater similarity between two exemplars.

	Blue	Green	Red	Near IR	IR1	IR2	FR
ROOT: mean	94.65	39.30	47.28	52.77	49.39	38.53	
std. dev.	20.63	12.73	23.54	23.94	27.00	26.48	1.1083
UR : mean	108.80	46.10	56.90	63.00	80.00	43.90	
std. dev.	5.59	2.28	3.96	5.44	3.23	2.28	1.1635
A1 : mean	80.90	34.10	28.80	34.20	75.60	21.80	
std. dev.	1.37	0.57	0.63	2.57	1.51	1.14	1.0904
A2 : mean	82.10	34.60	33.50	10.90	80.30	24.70	
std. dev.	0.99	0.70	0.71	2.23	2.16	1.25	1.0804
TG : mean	85.55	39.91	37.27	50.45	11.64	35.55	
std. dev.	1.51	1.51	1.49	3.17	2.20	1.63	1.0785
SD : mean	99.70	42.40	65.10	82.70	61.50	77.00	
std. dev.	1.64	0.84	1.66	1.16	3.50	1.89	1.1484
ND : mean	93.20	37.10	53.10	67.20	30.20	60.80	
std. dev.	1.69	0.99	1.37	1.32	3.08	1.40	1.1111
CO : mean	77.60	28.00	27.20	66.70	61.70	22.20	
std. dev.	1.07	0.94	1.75	3.77	4.81	3.26	1.1631
SW : mean	71.40	21.60	19.40	11.80	8.10	3.00	
std. dev.	1.78	0.97	0.97	1.03	1.52	1.15	1.1211
DW : mean	77.10	25.20	24.70	15.40	11.20	5.70	
std. dev.	2.02	0.92	1.34	0.70	1.14	0.95	1.0773
MA : mean	86.70	33.50	39.30	43.70	73.50	39.00	
std. dev.	2.16	1.35	2.54	2.54	4.84	2.94	1.1587
SS : mean	81.60	29.50	33.70	42.20	71.80	33.40	
std. dev.	1.51	0.53	0.67	0.92	1.14	1.17	1.0753
WS : mean	83.90	30.60	35.60	50.90	64.40	29.90	
std. dev.	1.29	0.52	1.07	2.23	2.59	1.37	1.1336
DB : mean	107.60	51.30	77.00	73.50	23.10	78.40	
std. dev.	2.59	2.11	3.62	4.74	4.72	3.44	1.1564
B1 : mean	123.54	60.77	75.46	85.00	61.69	31.31	
std. dev.	3.20	3.49	33.65	4.22	8.29	43.59	1.0824
B2 : mean	149.00	65.90	89.80	75.10	22.20	77.20	
std. dev.	4.67	1.52	2.30	1.52	2.53	2.04	1.1691

Figure 2: Standard deviations, means, and family resemblance scores for the 155-instance training set.

2.3 Classification and the family resemblance principle

When presented with a set of training instances, SX-WEB builds a three level tree structure. SX-WEB uses this tree structure together with its evaluation function to classify newly presented instances into one of the concept-level classes. When learning is not incremental, once an unknown instance is classified, it is discarded. In an incremental learning mode, the new instance becomes part of the classification tree. We now examine SX-WEB's evaluation function.

SX-WEB's evaluation function is based on the *family resemblance principle* (Cantor, 1979) which states that:

> Most prototypical members of a concept class share many features in common with members of their own class and few features in common with members of other closely related categories;

From a classification point of view, this principle implies that new instances to be classified should be placed in the category class that will result in a best overall family resemblance value as a result of instance inclusion. Based on this, we used a method proposed in (Tversky, 1977) for computing class family resemblance. Specifically:

$$FR(C) = 2/(N*(N-1)) * \Sigma \ Sim(a,b)$$

where C is the concept class whose family resemblance score is being computed, N is the total number of exemplars contained in concept class C, and Σ **Sim(a,b)** represents the sum total of all computed similarity scores between the class exemplars. In other words, to find the family resemblance score for concept class C, the similarity of each exemplar to all other exemplars in the class is summed. This sum is then divided by the total number of similarity computations made, giving an average similarity value for the class. Along these same lines, *typicality* is defined as the average similarity of one class exemplar to all other members of the class, or:

BAND	PRED VAL
BLUE	1.075
GREEN	1.134
RED	1.057
NEAR IR	1.192
IR 1	1.172
IR 2	1.055

Figure 3: Predictiveness scores for the attributes in the 155-instance training set.

$$Typ(e,C) = 1/N * \sum_{i}^{N} Sim(e,b_i) \text{ where } b_i e C \text{ and } b_i <> e$$

where e is the exemplar whose typicality score is being computed. The exemplar that gives a best score for the Typ function is known as the *class prototype* (Smith, 1981). The family resemblance statistic represents a global heuristic measure of classification goodness. Since similarity values closer to zero represent exemplars with greater similarity, an evaluation function that minimizes family resemblance scores would seem appropriate. Based on this idea, we define the evaluation rule used by SX-WEB:

> Given root node N, a list L of N's children representing the concept classes to be considered for instance classification, and a new instance I to be classified;
>
> Classify instance I with the concept class in L that results in the largest decrease in the average of the family resemblance scores of the children in L.

In other words, compute the average family resemblance score of all of N's children. Then take the first child C_1 in L and compute the new family resemblance score as a result of instance I being added to this child.

Now compute the new average family resemblance score resulting from this change to C_1. Subtract the new average family resemblance score from the old average. This is then the score for placing instance I into child C_1. This computation is made for each child in L.

Mathematically, since all that changes is the family resemblance score of the child now containing I and since the number of concept-level nodes remains constant, the actual computation is simply:

$$FR(C) - FR(C+I)$$

where C is the class in which I is being tested for incorporation and FR(C+I) is the family resemblance score of class C when I is included.

2.4 Complexity analysis

A cost analysis of SX-WEB's performance can be made for both the training and the classification component of the learning process. During the training phase, SX-WEB builds its tree by creating links between the root-level, the concept-level and the instance-level nodes. The root-level and concept-level nodes store sums and sums of squares rather than actual mean and standard deviation values to accommodate the possibility of an incremental learning environment.

After all of the training instances have been seen, the family resemblance score of the root-level and each concept-level node is computed. In a hierarchy containing R instance-level nodes, the family resemblance score of the root-level node can be computed by making R*(R-1)/2 similarity comparisons. If the R nodes are evenly distributed among M concept-level classes, then each concept-level class will contain approximately R/M instance-level children. This being the case, to find the family resemblance score for one concept-level class will require R/M*(R/M-1)/2 similarity computations. Therefore, to find the family resemblance scores for all M concept-level classes will require R*(R-M)/(2*M) similarity calculations.

The classification component cost analysis requires examining the total number of similarity computations necessary to classify a newly presented instance. To make instance classification as efficient as possible, each concept-level node stores a summation of instance similarity values rather than actual family resemblance scores. In this way, a new instance I being considered for classification into class C need only have its typicality score with the children of C computed. The typicality score for placing instance I into class C containing N children can be computed by making exactly N similarity computations. This typicality score can then be added to the present family resemblance summation value. From here the actual family resemblance score is computed by dividing the family resemblance summation by N*(N+1)/2. Therefore, to classify P instances using a concept hierarchy containing R instance-level nodes requires exactly P*R similarity computations.

When learning is incremental, each newly classified instance becomes an instance-level node within the concept hierarchy. In addition, the root-level node and the chosen concept-level node will have their statistics updated to reflect the incorporation of this new instance. An incremental learning environment is an advantage when concept class definitions need to be modified in order to reflect a changing learning environment.

When SX-WEB is used as an incremental learning system, classification efficiency changes significantly. This is true because each instance that becomes part of the learning hierarchy has the effect of modifying the standard deviation values of the attributes found within the chosen concept class. This results in the similarity values of all instances within the chosen class to be affected. Because of this, the incorporation of each new instance requires the family resemblance score for the chosen concept-level class to be recomputed. Specifically, in a hierarchy containing R instances and M concept-level classes where each class contains approximately R/M children, to determine which concept class will contain a newly presented instance I requires R similarity computations. Then, to update the family resemblance score for the chosen concept-level class requires approximately {R/M*[(R/M)+1]}/2 similarity computations.

3. EXPERIMENTAL RESULTS

This section gives the results obtained in testing SX-WEB using the derived Landsat TM data set previously described. The first three experiments test SX-WEB using all six spectral values. The remaining six experiments used predictiveness to test SX-WEB's classification accuracy when those spectral values least predictive of class membership were omitted from the classification process.

3.1 Classification utilizing all six spectral values

For the first experiment we used 155 of the 302 instances for the training phase. This resulted in a hierarchy containing fifteen concept-level nodes with each node representing one of the fifteen land cover categories. Individual concept-level nodes each contained ten training instances with the exception of the Turf/Grass (TG) class which contained eleven instances and the Barren 1 (B1) class with thirteen instances. The remaining 147 instances were then classified using SX-WEB's evaluation function. Of the 147 instances, 145 were classified correctly. One misclassification placed a Shrub Swamp (SS) instance into the Marsh (MA) class. The second misclassification placed a Southern Deciduous (SD) pixel into the Barren 1 (B1) concept class. The overall classification accuracy was 98.6%.

In the second experiment, we used sixty training instances, with four training instances being randomly chosen from each of the fifteen classes. Classification accuracy was 92.1%, with 223 of the 242 instances being correctly classified. When the 15 categories are generalized to seven categories [Urban (UR), Agricultural (AG), Deciduous (DE), Coniferous (CO), Water (WA), Wetland (WE), and Barren (BA)], so as to match the categories which were utilized in (Civco, 1992a), the accuracy rate increased to 96.2%, indicating that several of the misclassifications placed instances into similar concept categories.

In the third experiment the number of training pixels was reduced to 45 (3 randomly-selected pixels for each of the fifteen categories), and 257 pixels were then classified. Even with this small training set, 220 (85.6%) of the 257 pixels in the testing set were correctly classified. The specific incorrect classifications can be identified in the confusion matrix of Figure 4.

When the 15 categories are generalized to the seven categories of (Civco, 1992a), the accuracy rate improves slightly to 89.9% (231 of 257 pixels classified correctly). The resultant confusion matrix can be seen in Figure 5.

3.2 Classification utilizing less than six spectral values

Three experiments were performed using the 155 instance training set and different settings for the predictiveness threshold. In the first experiment the predictiveness threshold was set at 1.07 thereby eliminating the attribute RED from use during instance classification. (See figure 3 for predictiveness values.) When the remaining five attributes were used to classify 147 instances, the classification results were identical to those of experiment one (see subheading 3.1) giving an overall accuracy of 98.6%.

For the second experiment the predictiveness threshold was set at 1.13 thereby eliminating the spectral attributes BLUE, RED and IR2 from use. The results showed a classification accuracy of 96.6% with 142 of 147 instances being classified correctly.

In the third experiment, the predictiveness threshold was set at 1.17, which eliminated all attributes with the exception of NEAR IR from the classification process. The results obtained in using this single attribute for instance classification showed an overall classification accuracy of 55.8%.

Three additional experiments were performed using the 60 instance training set and various settings for the predictiveness threshold. The predictiveness values (not shown) for the 60 instance training set differed from those found in Figure 3. Specifically, BLUE was found to be the least predictive of class membership. NEAR IR and IR2 were the most predictive of class membership.

For the first experiment, a predictiveness threshold of 1.13 eliminated the attribute BLUE

	UR	A1	A2	TG	SD	ND	CO	SW	DW	MA	SS	WS	DB	B1	B2
UR	17														
A1		16	2		4								1	4	
A2		1	13												
TG			1	18				1		3	2	8			
SD					13										
ND						17									
CO			1				15								
SW								16							
DW							1		18						
MA										12					
SS										1	15				
WS										1		9			
DB													16	1	
B1														8	
B2							1							4	17

Figure 4: Confusion matrix, with columns representing actual categories of pixels and rows representing classifications by SX-WEB.

from use during classification. Instance classification resulted in 21 misclassifications and gave a 91.3% accuracy level.

In the second experiment, with a predictiveness threshold of 1.15, all attributes excepting BLUE and RED were predictive of class membership. The resulting classification showed 215 of 247 instances classified correctly giving an 89% accuracy level.

In the final experiment, a predictiveness threshold setting of 1.17 resulted in NEAR IR and IR2 being the only attributes predictive of class membership. Sixty two of the 242 instances were misclassified giving an accuracy rate of 74.4%.

3.3 Comparisons to other systems

As a means of comparison of these results to those obtained from other methods utilizing similar data sets, the reader is directed to (Civco, 1992a).

The results from (Civco, 1992a) can be partially summarized as follows:

The maximum likelihood estimation resulted in an overall classification accuracy of 91.5%.

A back-propagation neural network with a 6-element input layer, a 15-element hidden layer, and a 1-element output layer, resulted in an overall classification accuracy for 468 test pixels of 66.7%.

A similar network, but with both a 6-element hidden layer and a second 15-element hidden layer, resulted in an overall classification accuracy of 64.5%.

It is especially interesting to note that the greatest number of misclassifications by SX-WEB (see Figure 5) were the result of misclassifying Wetland (WE) pixels as Agricultural (AG) pixels. This misclassification was not present in the results found in the neural nets of (Civco, 1992a), although there was evidence of this type of misclassification with the maximum likelihood technique.

4. CONCLUSIONS AND FUTURE WORK

Recent research (Keil, 1987; Porter, 1990) supports an exemplar-based approach to concept learning. The findings of this research lends additional support to an exemplar-based concept learning paradigm. SX-WEB's similarity measure and evaluation function performed exceptionally well in the classification of pixel images representing fifteen different Landsat image types. High classification accuracy was achieved even when each concept class contained as few as three training instances. The results of predictiveness testing were also positive in that high levels of classification accuracy were maintained when a limited number of

	UR	AG	DE	CO	WA	WE	BA
UR	17						
AG		51	4		1	13	5
DE			30				
CO		1		15			
WA				1	34		
WE						38	
BA				1			46

Figure 5: Confusion matrix for generalized categories, with columns representing actual categories of pixels and rows representing classifications by SX-WEB.

spectral values were used for the classification process.

SX-WEB is currently running on an Intel (TM) 80386-based machine with a clock speed of 33 mhz and without a math coprocessor. The times needed to run the experiments specified in this paper were 10 to 35 minutes. This was largely dependent on the size of the training set. It is assumed that running SX-WEB on an Intel (TM) 80486-based machine with a higher clock speed would significantly improve performance, resulting from both the increased clock speed and the integrated math coprocessor. Once this type of machine is available to the authors, a much larger data set will be used to empirically evaluate the time requirements of the system.

In addition, PC Scheme has a fairly high overhead for "garbage collection," and it should be possible to rewrite the program to minimize this, or to implement SX-WEB in another language, such as C.

The authors are currently preparing a data set from Landsat MSS data acquired over southern Minnesota in July of 1988. The limitation of the input to SX-WEB to the four spectral bands (0.5-0.6 μm, 0.6-.07 μm, 0.7-0.8 μm, 0.8-1.1 μm) should be instructive.

Another area of endeavor will be to utilize EX-WEB's abilities to perform incremental learning and unsupervised classification. SX-WEB will be trained to perform classification into two categories (Water/Wetland and Other), using a small subset of the southern Minnesota data set. The resultant classification tree will then be used to extract all Water/Wetland pixels from the entire data set. These pixels will then be the input for an unsupervised incremental classification using EX-WEB. This will done in order to further differentiate between Water/Wetland types.

This classification of Water/Wetland types is currently being performed manually in the Water Resources Center at Mankato State University, and it would appear that automation may be possible.

REFERENCES

Cantor, N. and Mischel, W. 1979. Prototypes in person perception. Advances in Experimental Social Psychology, 12:3-52.

Civco, D.L. 1991. Landsat TM land use and land cover mapping using an artificial neural network. In: Proc. of the 1991 Annual Meeting of the American Society for Photogrammetry and Remote Sensing, Baltimore, MD. 3:67-77.

Civco, D.L. 1992a. AI and the environment. PC AI, 6(1):43-45,56.

Civco, D.L. 1992b. Artificial neural networks for land cover classification and mapping. International Journal of Geographic Information Systems. (in press)

Keil, F.C. 1987. Conceptual development and category structure. In: U. Neisser (Ed.), Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization. Cambridge University Press, New York.

Lillesand, T.M. and Kiefer, R.W. 1987. Remote Sensing and Image Interpretation, Second Edition. John Wiley and Sons, New York.

Porter, B.W., Bareiss R. and Holte R.C. 1990. Concept learning and heuristic classification in weak-theory domains. Artificial Intelligence, 45:229-263.

Roiger, R.J. 1991. Incremental Concept Formation In An Instructional Environment. Unpublished doctoral dissertation, University of Minnesota.

Rosch, E., and Mervis, C.B. 1975. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology: 573-603.

Rosch, E., Mervis, C.B., Gray, W., Johnson, D., and Boyes-Braem, P. 1976. Basic objects in natural categories, Cognitive Psychology, 8:382-439.

Smith, E.E., and Medin, D.L. 1981. Categories and Concepts. Harvard University Press, Cambridge, MA.

Tversky, A. 1977. Features of similarity, Psychological Review 84:327-352.