Automatic Discovery of Classification and Estimation Algorithms for Earth-Observation Satellite Imagery

Rick L. Riolo
Program for Study of Complex Systems (PSCS)
University of Michigan
Ann Arbor, Michigan
rlriolo@umich.edu

Mark P. Line
Control Data Systems, Inc.
Seattle, Washington
Mark.P.Line@cdc.com

Abstract

Under NASA's new Earth Observing System (EOS), satellite imagery is expected to arrive back on Earth at rates of gigabytes/day. Techniques for the extraction of useful information from such massive data streams must be efficient and scalable in order to survive in petabyte archive situations, and they must overcome the opacity inherent in the data by classifying or estimating pixels according to user-specified categories such as crop-type or forest health.

We are in the process of applying GP to several related satellite remote sensing (RS) classification and estimation problems in such a way as to surmount the usual obstacles to large-scale exploitation of imagery. The fitness functions used for training are based on how well the discovered programs perform on a set of cases from Landsat Thematic Mapper (TM) imagery. Programs are rated on how well they perform on out-of-training-set samples of cases from the same imagery.

We have carried out a number of preliminary experiments on a relatively simple binary classification task. Each case is a set of 7 spectral intensity readings for a pixel and an associated ground truth class: 1 for surface water, 0 for none. The GP system very rapidly discovers simple relations that correctly predict 98%+ for training and testing data sets. The key problem with the results we have observed so far is that the simple solutions rapidly drive out diversity in the population. Several approaches will be taken in further study in order to try to maintain diversity in the population.

Problem Description

A current hotbed of research in the satellite remote-sensing community involves the problem of extracting useful information from digital imagery and making that information as generally accessible as possible – as exemplified by the recent NASA Cooperative Agreement Notice (CAN), "Public Use of Earth and Space Science Data Over the Internet." Any approach to solving this problem must eventually surmount two major obstacles: data volume and data opacity.

Under NASA's new Earth Observing System (EOS) (Congress, 1994), satellite imagery is expected to arrive back on Earth at the rate of 220 GB/day; the data volumes to be processed through and managed by the EOS information system are measured in petabytes. Most methods for the extraction of useful information from image data can be categorized as classification techniques or as estimation techniques (Aronoff, 1989; Ch. 3). The goal of classification techniques is to assign correctly (i.e. with known accuracy) each (potentially multibyte, multispectral) image pixel to one of a finite set of (not necessarily previously known) classes, such as land-cover classes (water, forest, crops, rock, ...). The aim of image estimation techniques is to calculate correctly (with known accuracy) the relative proportions of each class in the image, but with no commitment to the correct classification of any given pixel taken alone. Classification and estimation techniques to be deployed by EOS and similar systems will have to confront efficiency and scalability issues as a first priority.

The second major obstacle to extracting useful information from satellite imagery is the inherent opacity of the data: direct visualizations of multispectral imagery are practically meaningless to users not trained in their interpretation.

Frameworks for the development (or automatic discovery) of classification and estimation algorithms for satellite imagery must be examined in the light of these obstacles – if they are to be evaluated as to their long-term viability. Specifically, the framework must deliver methods which are

- easy to create using supervised learning without expert knowledge of remote sensing or machine learning;
- easy to understand without cumbersome and time-consuming visualization techniques;
- easy to apply to fresh imagery data without expert knowledge of remote sensing or machine learning; and
- computationally efficient and scalable to petabyte-sized archives of data objects.

We suggest Genetic Programming (GP) as a framework which stands to fare favorably under each of these

criteria as a long-term solution to the information extraction problem, and we have begun to explore such applications of GP in the research-in-progress reported here.

Previous Work

Three basic paradigms have been applied to the problem of image classification and estimation in remote sensing:

- "classical" statistics-based techniques, e.g. (Schowengerdt, 1983);
- knowledge-based approaches (Kartikeyan et al., 1995) (Goodenough et al., 1994) (Ton et al., 1991) (Wharton, 1987); and
- supervised learning with artificial neural networks (ANN's) (Wu and Westervelt, 1994) (Civco, 1993) (Bischof et al., 1992) (Heermann and Khazenie, 1992).

The development of classification and estimation methods based on classical statistical techniques is inherently very difficult – these techniques are only accessible to highly-trained and experienced researchers. A high level of expertise is often required to interpret the results correctly, as well. These disadvantages conspire to reduce greatly the potential long-term viability of these techniques for general use in the environment described above.

Although the use of knowledge-based techniques can be very helpful in organizing large corpora of interrelated classification and estimation methods, these techniques actually exacerbate the disadvantages of statistics-based techniques in that expertise is required not only in the underlying image-analysis technology, but also in the area of knowledge engineering.

A similar objection applies to the use of ANN-based techniques in remote sensing: due to the level of human intervention currently required to cajole ANN's into training appropriately in a given problem domain, the researcher or research team must bring expertise to bear not only from remote sensing but also from neurocomputing. In addition, both successfully and unsuccessfully trained ANN's are inherently difficult to understand and interpret – even for highly-qualified experts in the techniques employed. Although high-level graphical visualization tools may aid researchers in examining the patterns of synaptic weights derived during training, the relationship between these patterns and recognizable phenomena in the problem domain is anything but straightforward.

Rationale of GP Approach

Genetic Programming (GP) uses the ideas of natural selection to create computer programs that solve user-specified problems (Koza, 1992). GP has been applied to a wide variety of standard machine-learning problems, from robot control to time-series prediction (Koza, 1992, 1994a) (Kinnear, 1994a). In the last couple of years, GP has begun to be applied to "real world" problems, e.g. in the classification of amino acids sequence domains (Koza, 1994b). Of particular relevance to Remote Sensing (RS), GP has been used in binary classification of objects based on features extracted from IR images of landscapes (Tackett, 1993). And most recently, GP has been applied to extracting features from satellite images (Daida et al., to appear).

The results obtained with GP so far, especially the recent results on very difficult problems, recommend GP as a technique for discovering algorithms to extract information from satelite imagery. GP seems well suited to solving the information extraction problems outlined in Section 1, to wit:

- Good results often can be obtained without relying on detailed domain knowledge supplied by experts; instead, sufficiently good classifiers for new themes can be created rapidly using supervised learning on a supply of training cases.
- The results are usually readily interpretable, and they are easily transformed into efficient implementations in conventional image-processing platforms.
- When large volumes of data must be used to achieve high accuracy, GP can readily be run in parallel, with large populations and numbers of generations (Koza, 1994b).

Furthermore, GP offers several additional advantages over the standard techniques described above:

- The same basic GP techniques might be used to achieve different accuracy requirements. For example, GP could be used to rapidly discover an estimator or classifier with 80% accuracy, which for many situations is all that is required (e.g. an estimator for percent crop cover, or a classifier for data-mining in image archives). For other situations, more GP resources could be used to discover much more accurate classifiers (e.g. for thematic mapping of land-cover categories).
- Results might be transferrable from one task to other related ones, e.g. in the form of reusable Automatic Defined Functions (ADFs) (Koza, 1994a).

For these reasons we are currently using GP to automatically discover algorithms to extract information from satelite imagery, as described in the next section.

Current and Projected Work

We are in the process of applying GP to several related RS classification and estimation problems. For both estimation and classification, there are two types of problems we are working with initially, based on whether there are two or more classes to be distinguished:

- binary classes: for example, classify each pixel as representing either water or not-water; or give an estimate of the percentage of water in an image

- nominal classes: for example, classify each pixel as to whether it represents forest, field, water, bare-ground, wetlands, or urban territory

For all problems, the basic plan is to use GP to automatically discover algorithms which solve the classification or estimation task. The fitness functions used for training will be based on how well the discovered programs perform on a set of cases from Landsat Thematic Mapper imagery (i.e., we will be doing supervised learning). For evaluation and comparison of the results to those obtained by other techniques (or by other GP parameter settings), programs will be rated on how well they perform on out-of-training-set samples of cases from the same imagery.

The imagery data consists of pixel-by-pixel intensity values for 7 spectral bands (4 visible, 3 near and mid infrared, 1 far (thermal) infrared). Each pixel corresponds to a 30m by 30m area on the ground. The 7 bytes of band data for each pixel in the training set are joined by a ground-truth value representing the information to be extracted from the image (e.g. land-cover class). Thus, given the availability of ground-truth data for the same area as that covered by the Landsat image, it is possible to generate literally millions of fitness cases for training and testing sets: a single Landsat TM image contains about 36 million pixels, corresponding to a 185km by 185km area on the ground.

We plan to use data sets of varying difficulty, including:

- data sets constructed from real data, in which we manipulate the complexity of solutions required, the amount of noise in the data, the amount of category overlap, and so on;
- real data sets with known optimal solutions; and
- real data sets with unknown optimal solutions.

By using data sets with known solutions and complexity, we will be able to systematically test and compare the results obtained using different modifications of the basic GP approach, as in (Tackett and Carmi, 1994).

Preliminary Results and Discussion

We have carried out a number of preliminary experiments on a relatively simple binary classification task for which a near-optimal solution is known. The data sets were extracted from a segment of recent Landsat TM imagery covering about 1,000 square kilometers of West Central Louisiana. Each case is a set of seven spectral-band intensity readings (integers from 0 to 255) and an associated "ground truth" class: 1 for water or 0 for not water. Thus the goal is to find rules (programs) that can predict whether the pixel represents water or not, given the intensity readings at the various spectral bands available.

Since the shape of the spectral histogram of sunlight reflecting off of surface water through the Earth's atmosphere is known (decreasing monotonically in intensity from blue through red to the near and mid infrared), the presence of surface water is conventionally detected by merely checking for such a monotonic decrease – this method is robust for different absolute intensities (brighter days) and for large variations in inter-band slopes (muddier water). In Landsat-TM imagery, the bands resulting from multispectral scanning are as follows:

1: blue (0.45-0.52 microns)
2: green (0.52-0.60 microns)
3: red (0.63-0.69 microns)
4: near-infrared (0.76-0.90 microns)
5: mid-infrared (1.55-1.75 microns)
6: thermal-infrared (10.4-12.5 microns)
7: mid-infrared (2.08-2.35 microns)

Therefore, the rule for detecting surface-water from bands 1 through 7 can be expressed as

$$(B_1 > B_2 > B_3 > B_4 > B_5 > B_7)$$

where B_i is the intensity-value of the i'th band. Note that thermal-IR intensities are irrelevant to this rule. Also note that for some particular data sets, a subset of this rule, e.g, $(B_3 > B_4)$, is sufficient to correctly categorize a high percentage of the points.

To solve this using GP, we used a terminal set

$$T = \{B_1\ B_2\ B_3\ B_4\ B_5\ B_6\ B_7\ R\}$$

which supplies values at each band and random ephemeral constants in the range [-7,7]. The function set used was

$$F = \{+\ -\ /\ *\ <\ >\ =\ AND\ OR\}.$$

We have tried two versions: one with mixed types and one with all integer types, each with the appropriate wrappers, operator protection and result conversions. Typical population sizes were 200 to 500, and maximum generations run was 50. We have tried two raw fitness functions: %Corrrect, and a correlation measure C which runs from -1 (completely wrong) through zero (random guesses) to 1 (complete correct) (Koza, 1994). Each was mapped into a standardized fitness running from 0 to 1.

In the runs we have done so far, the GP system very rapidly discovers simple relations (e.g., $B_3 > B_4$) that correctly predict 98%+ for training and testing data sets. These results were obtained using both the mixed-type and integer versions, and for several data sets. This is very encouraging, even given the simplicity of the solutions found, in part because it is unlikely that an Artificial Neural Net approach would have obtained such good performance as rapidly, and of course the ANN solution would not be as clearly interpretable as the GP programs we obtained (being directly referrable to known reflectance properties as they are).

As mentioned earlier, rules that can correctly classify 85% or more of the cases are often all that is required; being able to find such rules rapidly and robustly is exactly one of the results we hope to show GP can produce.

However, for some tasks it is necessary to get very high percent-correct rates, e.g. for thematic mapping. For these tasks, we will want the GP to go beyond the simple but pretty good solutions, to discover more complex solutions that get that last few percentage of cases correct as well.

The key problem with the results we have observed so far is that the simple solutions rapidly drive out diversity in the population. In particular, the logical operators are completely lost from the population. Once that happens, there is no way for the GP to create plausible candidate solutions by recombining existing solutions. This loss of diversity was observed using both fitness proportionate (FP) selection and binary tournament selection (TS). However, we did note that diversity was maintained longer using TS, as would be expected given its reduced selection pressure.

These results highlight what will be a key issue for tasks in which the goal is to find a very high-performance classifier: how to maintain population diversity so that the system will not get stuck on easily-found local optima, but instead continue on to find solutions that require more complex programs involving all the basic functions and feature values.

Over the next few months, we plan to try a number of approaches to maintaining diversity in populations. In particular we plan to explore:

- controlling selection pressure, e.g. by scaling fitness values or by using a lower threshold for lower fitness individuals to win in binary TS;
- subpopulation mixing approaches (demes), e.g. as used in (Tackett and Carmi, 1994) or in (Koza, 1994a); and
- coevolution of fitness cases, to give more weight to those cases that are not solved by simple expressions, e.g. as done in (Siegel, 1994).

We may also explore the use of special operators besides the standard subtree-swapping crossover. For example, it may be useful to have a recombination operator that combines subtrees by introducing logical operators to join them. For example, if one individual has discovered $(> B_3\ B_4)$ and another (perhaps in a separate subpopulation) has discovered $(> B_2\ B_3)$, the operator could combine these using AND to form $(AND\ (> B_3\ B_4)\ (> B_2\ B_3))$, which is a step on the way toward complex relations like $(B_1 > B_2 > B_3 > B_4 > B_5)$. Similar operators could introduce other logical or conditional connectives.

We also plan to explore the use of various types of Automatically Defined Functions (Kinnear, 1994b). For example, ADFs will probably be useful for discovering systematic operations that can be applied to a pixel and its neighbors, to include information about a pixel's context in deciding how to classify it. However, that work will not be the focus of this paper.

Additional Results

A longer version of this paper, including all the results discussed at the AAAI Fall Workshop, can be obtained via anonymous ftp from pscs.physics.lsa.umich.edu/papers/95/gp1-long.tar.gz, which includes both postscript and dvi copies.

Acknowledgements

Riolo was supported in part by the UM Program for Study of Complex Systems and by National Science Foundation Grant IRI-9224912. Line was supported in part by the US Army Corps of Engineers Construction Engineering Research Laboratories (USACERL) and by the National Center for Supercomputing Applications (NCSA).

References

Aronoff, Stan. 1989. Geographic Information Systems: A Management Perspective. Ottawa, Ontario: WDL Publications.

Bischof, H., W. Schneider and A.J. Pinz. 1992. "Multispectral classification of Landsat-images using neural networks," *IEEE Transactions on Geoscience and Remote Sensing* 30(3): 482-490.

Civco, Daniel L. 1993. "Artificial neural networks for land-cover classification and mapping," *International Journal of Geographical Information Systems* 7(2): 173-186.

Congress, Office of Technology Assessment. 1994. Remotely Sensed Data: Technology, Management, and Markets. OTA-ISS-604. Washington, D.C.: US Government Printing Office.

Daida, J.M., J.D. Hommes, S.J. Ross and J.F. Vesecky (to appear). "Extracting curvilinear features from SAR images of arctic ice: Algorithm discovery using the genetic programming paradigm," to appear in Proceedings of IEEE International Geoscience and Remote Sensing 1995, Florence, Italy.

Goodenough, D.G., D. Charlebois, S. Matwin and M. Robson. 1994. "Automating reuse of software for expert system analysis of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing* 32: 525-533.

Heermann, Philip D. and Nahid Khazenie. 1992. "Classification of multispectral remote sensing data using a back-propagation neural network," *IEEE Transactions on Geoscience and Remote Sensing* 30(1): 81-88.

Kartikeyan, B., K.L. Majumder and A.R. Dasgupta. 1995. "An expert system for land cover classification," *IEEE Transactions on Geoscience and Remote Sensing* 33(1): 58-66.

Kinnear, Kenneth E. Jr., ed. 1994a. Advances in Genetic Programming. Cambridge, MA: MIT Press (Bradford Books).

Kinnear, Kenneth E. Jr. 1994b. "Alternatives in automatic function definition: A comparison of performance," in Kenneth E. Kinnear, Jr. (ed.), Advances in Genetic Programming. Cambridge MA: MIT Press (Bradford Books).

Koza, John. 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, Cambridge MA.

Koza, John. 1994a. Genetic Programming II: Automatic Discovery of Reusable Programs. Cambridge MA: MIT Press.

Koza, John. 1994b. "Evolution of a computer program for classifying protein segments as transmembrane domains using genetic programming," in R. Altman et al. (eds.), Proceedings of Second Inter. Conf. on Intelligent Systems for Molecular Biology.

Schowengerdt, Robert A. 1983. Techniques for Image Processing and Classification in Remote Sensing. NY: Academic Press.

Siegel, Eric V. 1994. "Competitively Evolving Decision Trees Against Fixed Training Cases for Natural Language Processing," in Kenneth E. Kinnear, Jr. (ed.), Advances in Genetic Programming. Cambridge MA: MIT Press (Bradford Books).

Tackett, Walter Alden. 1993. "Genetic programming for feature discovery and image discrimination," in Stephanie Forrest (ed.), Proceedings of the Fifth International Conference on Genetic Algorithms. San Mateo CA: Morgan Kaufmann.

Tackett, Walter Alden and Aviram Carmi. 1994. "The Donut problem: Scalability, generalization and breeding policies in genetic programming," in Kenneth E. Kinnear, Jr. (ed.), Advances in Genetic Programming. Cambridge MA: MIT Press (Bradford Books).

Ton, Jezching, Jon Sticklen and Anil K. Jain. 1991. "Knowledge-based segmentation of Landsat images," *IEEE Transactions on Geoscience and Remote Sensing* 29(2): 222-231.

Wharton, S.W. 1987. "A spectral-knowledge-based approach for urban land cover discrimination," *IEEE Transactions on Geoscience and Remote Sensing* 25(3): 272-282.

Wu, Xiping and James D. Westervelt. 1994. Using Neural Networks to Correlate Satellite Imagery and Ground-Truth Data. US Army Corps of Engineers, USACERL Special Report EC-94/28. (available from the National Technical Information Service, 5285 Port Royal Road, Springfield VA 22161)