N82-13723 D58

STRUCTURE OF THE KNOWLEDGE BASE FOR AN EXPERT LABELING SYSTEM*

By N. S. Rajaram

Lockheed Engineering and Management Services Company, Inc.

SUMMARY

One of the principal objectives of the NASA AgRISTARS program is the inventory of global crop resources using remotely sensed data gathered by Land Satellites (Landsat). A central problem in any such crop inventory procedure is the interpretation of Landsat images and identification of parts of each image which are covered by a particular crop of interest. This task of "labeling" is largely a manual one done by trained human analysts and consequently presents obstacles to the development of totally automated crop inventory systems. However, development in Knowledge Engineering as well as widespread availability of inexpensive hardware and software for Artificial Intelligence work offers possibilities for developing expert systems for labeling of crops. Such a knowledge based approach to labeling is presented in this paper.

INTRODUCTION

The Landsat spans different parts of the earth's surface providing images at regular periodic intervals. Images gathered by Landsat are in the form of picture elements (or pixels) consisting of the average spectral response of the area covered by each pixel (about an acre) in four visible and near infrared frequency bands. It is known however that it is possible to make a transformation of this four-channel data onto a two-dimensional plane essentially preserving the information contained in the original data. This permits creation of visual displays to be used by analyst interpreters (labelers) for assigning labels to different parts of the scene. (cf: Ref. [2], [4], and [6]). It is to be noted that labeling forms only a part of the overall remote sensing crop inventory exercise, but as experience has shown, a most crucial part. It may also be mentioned that labeling of images arises in other applications besides crop inventory by remote sensing (Ref. [9]).

In all decision making processes including labeling, human experts use knowledge which is not easy to formalize. Such expert knowledge can frequently consist of neuristic rules, elimination by constraints, exercise

*Work for this project being performed for the NASA Earth Resources Division under Contract NAS-9-15800 at the Lyndon B. Johnson Space Center.

of judgment, all gained through experience. In the context of labeling, experienced analysts make successful use of spatial context and texture. It is precisely these skills which one would like to impart to an automatic labeler.

One of the most intriguing concepts in automation technologies is the idea of employing Artificial Intelligence (AI) systems in a decision making capacity, in effect replace human experts by "Expert Systems." One of the most active areas of AI research is the application of such techniques to problems in Pattern Recognition/Image Understanding (Ref. [8] and [9]). This development is hardly a coincidence. AI and Pattern Recognition share several common features among which the most significant is that they are "knowledge" based (Ref. [8]). A fundamental requirement in AI approaches is that one be able to represent and process "knowledge" and not merely data. We briefly discuss a knowledge based approach to labeling in remainder of this paper.

KNOWLEDGE BASED APPROACHES

Knowledge Engineering is the branch of Artificial Intelligence used in building expert systems. In it one attempts to capture the essential problem solving skills of an expert, transmit those same skills to a computing system thereby creating an automated expert or an expert system. It is now recognized (Ref. [2]) that human problem solvers possess knowledge and techniques which are specific to a problem area and not general problem solving skills. Furthermore, this knowledge is frequently heuristic knowledge consisting of judgment, experience, good practice, and so on. Thus, it is clear that any expert system has to include a store of knowledge called "Knowledge Base" and a set of techniques usually called "Paradigms." Consequently, an effective and flexible representation of knowledge is a crucial first step in attempting to develop any expert system.

The pioneering effort in knowledge based problem solving was undoubtedly Slagle's integration program (Ref. [8]). Since that time many expert problem solvers ranging in applications from advanced mathematics to medical diagnosis have been successfully built. Of these we can cite MACSYMA (advanced mathematics), MYCIN (diagnosis of blood disorders), and DENDRAL (chemical analysis of spectroscopic data) as examples of highly successful knowledge based expert systems. As a result of these efforts, the following general insights have been gained.

1. Any Artificial Intelligence work entails storage and manipulation of complex data structures.

2. These complex data structures which represent "knowledge" can include data as well as programs. Hence, the programming environment must facilitate uniform representation of data and programs.

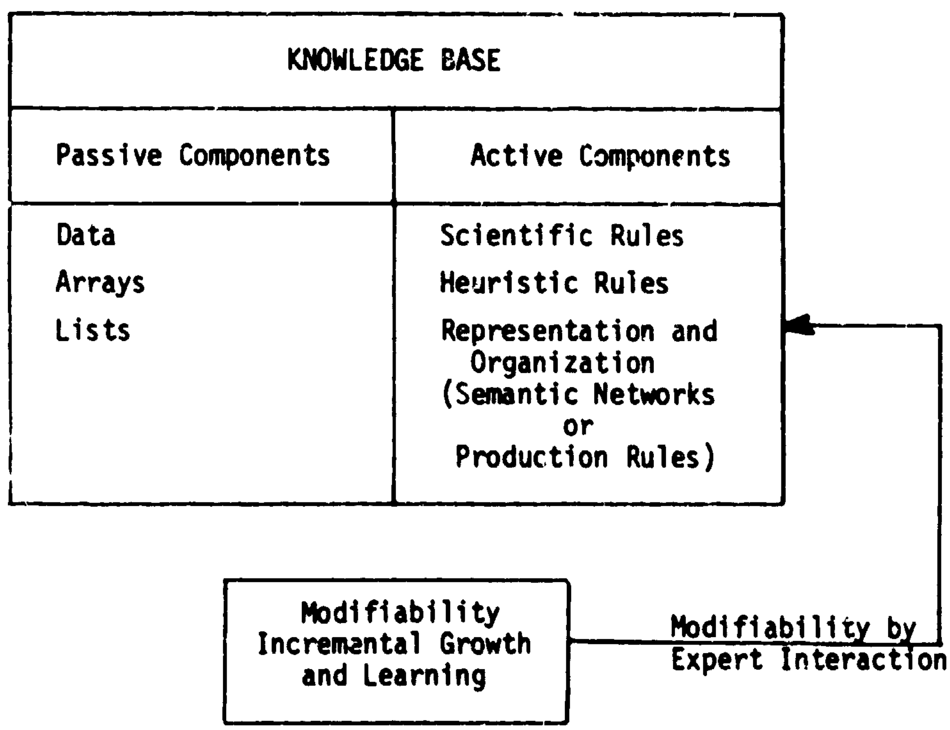
3. This activity involves "knowledge processing" and not just "data processing."

4. Such a knowledge based system should permit ease of frequent modification as we update our knowledge, without dismantling the system. This is known technically as "embedability."

5. These conditions argue for a LISP base for AI systems. In fact LISP is the _sine qua non_ for serious AI work. (Ref. [8])

In addition to its uniform treatment of data and programs as symbolic expressions, LISP as a programming language provides several extremely powerful features such as lambdas and recursive function calls for transfer of control and use of programs and functions as arguments for other programs.

We can represent a knowledge base by the following general schematic diagram.



KNOWLEDGE BASE FOR THE EXPERT LABELING SYSTEM

Recall that we are interested in assigning different parts of a scene to different land cover types based on its Landsat imagery or "label" the scene. We also noted that the 4-dimensional data obtained by the multispectral scanner (MSS) are transformed into a two-dimensional space known as the "greenness-brightness" space. Consequently, data to be used for labeling will consist of multitemporal greenness-brightness values for each pixel. We shall see later that there are certain minimal data requirements. The production rules will be based on a study conducted by Palmer and Magness (Ref. [5]), and contain a very large heuristic base. These rules will be displayed in a widely used version of LISP known as "MACLISP" (Ref. [9]). We shall not discuss the labeling paradigms separately but merely note that inference rules with the help of IF THEN--- rules by COND constructs present no particular difficulties (Ref. [9]).

Because of atmospheric conditions not all the Landsat passes over a scene result in usable data acquisitions. Such deficiencies result in imperfect or incomplete labeling. For the sake of exposition, we assume that we are attempting a two-stage decision, assigning a label as Spring Small grains or not at stage 1 and as barley or not at the final stage. In order for a pixel to be labelable, the minimal acquision requirements are expressed in terms four "windows" which are time intervals in which data are available. Based on Palmer-Magness study, we define the four "windows" which are intervals represented by lists as follows:

```
(SETQ W1 (LIST

(DIFFERENCE X1 5) (PLUS X1 18)))
```

Where X1 is computed from the crop calendar time when at least 50 percent has been planted for wheat. Based on similar considerations, windows W_2 for wheat W_3 and W_4 for barley are defined. These values for X1, et al are determined from models for crop calendars. Essentially they ascertain availability of observations during certain crucial crop development stages. It is an interesting empirical observation that minimum data requirements are different for different geographical regions. This "contextual information" forms a crucial part of the knowledge base. Next, we check to see if the Landsat data are adequate for labeling. For each pixel, there is a list of acquisition dates from which we decide whether or not the region of that pixel is labelable. The following function checks if the list DATA contains an acquisition in particular window by recursive scanning the list DATA.

```
(COND ((NULL DATA) 'ABSENT)

      ((GREATERP (CAR WINDOW)

                 (CAR DATA)
```

```
                    (CADR WINDOW)) 'PRESENT)
              (T (CHECK WINDOW (CDR DATA)))))
```

For example, the function (CHECK W1 DATA) will return PRESENT or ABSENT whether or not DATA contains an acquisition in the window W1. This function is evaluated for each of the four windows. Then it is a simple matter to write LISP function which determines whether the minimal data requirements are met, for that particular geographical region. In addition, the knowledge base includes information concerning quality of the data in the form of cloud cover, haze, etc. If such cover exceeds 40 percent, the scene cannot be labeled. The actual labeling procedure or "Paradigm" uses a scatter plot generated from windows 2 and 3 and a decision line which will not be discussed here. It is to be noted that we have given only the outline of the procedure omitting most of the details. An alternative approach, based on an angular statistical measure to crop classification can be found in Ref. [7].

So the knowledge base for an expert system based on the scheme described above can be envisaged to be built around the following skeleton.

1. An eliminating constraint which discards data with excessive haze and/or cloud cover.

2. A production rule which determines the four windows as described earlier. (The window function.)

3. A decision rule which checks the geographical location of the scene against the minimum data requirements in the form of acquisitions in the windows. (The function CHECK.)

4. Data will comprise of crop calendar information as well as Landsat data.

It is interesting to note with the exception of the data, all the rules given above use programs or functions as components in the knowledge base.

Acknowledgments: The author would like to thank his collegues R. P. Heydorn of NASA, R. K. Lennington, E. R. Magness and T. C. Minter of Lockheed and D. Y. Y. Yun of I.B.M.

REFERENCES

[1] R. O. Dudla, P. E. Hart, N. J. Nilsson, and G. L. Sutherland (1978) (Semantic Network Representations in Rule-Based Inference Systems," in Pattern Directed Inference Systems, edited by D. A. Waterman and F. Hayes-Roth

[2] E. A. Feigenbaum (1978) "The Art of Artificial Intelligence — Case Studies in Knowledge Engineering." Proceedings of the National Computer Conference.

[3] R. B. MacDonald and F. G. Hall (1980) "Global Crop Forecasting." Science, Vol. 208, pp 670-679.

[4] A. Newell and H. A. Simon (1972) Human Problem Solving. Prentice Hall, Englewood Cliffs, New Jersey.

[5] W. F. Palmer and E. R. Magness (1981) A Description of the Reformatted Spring Small Grains Labeling Procedure. NASA-Lyndon B. Johnson Space Center Report JSC-16827.

[6] C. V. Srinivasan (1980) Knowledge Based Learning Systems Department of Computer Science. Report DCS-TR-89, Rutgers University, New Brunswick, N J.

[7] S. G. Wheeler and P. N. Misra (1980) "Crop Classification with Landsat Mutlispectral Scanner Data II." Pattern Recognition, Vol. 12, pp 219-228.

[8] P. H. Winston (1977) Artificial Intelligence, Addison-Wesley, Reading Massachusetts.

[9] P. H. Winston and B. K. P. Horn (1981) "Expert Problem Solving Using IF — THEN Rules," pp. 239-250 in LISP, Addison-Wesley, Reading, Massachusetts.

1830 NASA Road 1
Houston, Texas 77058, U.S.A.