PROGRESS IN KNOWLEDGE ENGINEERING FOR IMAGE INTERPRETATION AND CLASSIFICATION.

N.J.Mulder, H.Middelkoop, J.Miltenburg.
I.T.C.
P.O.Box 6, Enschede.
The Netherlands.

III/4

ABSTRACT :

The common grounds between pattern recognition, geo information systems and expert systems . The "expert system" is de-mystified, useful techniques in knowledge engineering are identified. The effects of the degeneration of the method of statistical pattern recognition into a rulebase are studied. The development of an "expert tutor" for the interpretation of soil classes from SPOT images is reported. Implementations are based on a home developed PROLOG based shell with binary logic and a commercial shell (ENVISAGE) providing limited reasoning with uncertainties.

INTRODUCTION

Expert systems and artificial intelligence are popular / fashionable subjects probably because of mans fascination with building things after the likeness of his own image. From this driving force in the computing community a number of software packages have resulted, mostly in the field of man - machine interfacing . If we first remove the mystic layers covering the lines of code then we discover techniques for knowledge engineering which are based on finding and recording traces in a space (matrix) which relates the instantiation of a problem to the instantiation of a solution.

At ITC we are mostly interested in the engineering of knowledge used in classification and interpretation. When expert systems are refered to in the rest of the paper those will be assumed to be of the diagnosis type.

Work on the use of knowledge engineering in the fields of image interpretation and mapping at I.T.C. was initiated in 1980, resulting in publications in 1982,(1984) on context dependent classification using terrain knowledge in a selective way.

A first meta rule seemed to be : *define a hierachy of decision rules , use data selectively .*

In 1985 v.d.Pluym and Mulder (not published) investigated the relation between rules in expert systems and rules in pattern recognition. This investigation concentrated on the role of certainty factors. It was soon apparent that the certainty factors used in medical expert systems were not suitable because fuzzy logic does not optimise anything. The Bayes decision rule on the other hand, when combined with a cost / benefit matrix maximises the (economical , social, political) benefit of the application of the classification rule.

So the second meta rule appears to be :

base certainty factors on likelyhood estimation, bound to the benefit of proper classification and cost of wrong classification .

The objection against statistical certainty factors in medical expert systems is " there are not enough samples available to define the multi variate statistical distributions"

. In remote sensing and mapping this is not a problem and moreover an intuitive estimation of statistical relations is more apt to provide a good parameter than the ill defined measures of fuzzy logic..

In order to make some of the knowledge engineering techniques clear to staff and students several versions of an "expert" system have been developed at ITC by the authors based on micro PROLOG and ENVISAGE.

DE-MYSTIFICATION OF THE EXPERT or CAN KNOWLEDGE BE COMPILED ?

As the decision rules map the observation vectors in a unique way into class likelyhood vectors, and as the maximum likelyhood class, $P_{max}(class|X)$ is choosen ; the relation between, all possible inputs (X) and outputs (class), is fixed and could be compiled in one large decision (truth) table .

An important part of the de-mystification of expert systems and artificial intelligence is the realisation that any expert system could be replaced by a large enough library with a sophisticated indexing system. This is demonstrated in one of the PROLOG based versions. (In pattern recognition use is made of classification look-up -tables).

CASE STUDY ; SOIL MAPPING FROM SATELLITE IMAGERY.

Chemical and textural characteristics of a soil profile, which are used in soil classification systems, such as the US Soil Taxonomy (SOIL SURVEY STAFF,1975), cannot be observed from space. Therefore it is nescessary to define the relation between soil class and landscape features, e.g. types of vegetation, altitude, shape of natural landsegments.

After a first inventarisation of the world soil types and possible relevant landscape features it turned out that this relation strongly depends on local (agri) cultural interferences in the past and present. This is contrary to the more definition based, rules for the ordinary classification of soil profiles. Therefore ,a large amount of local contextual knowledge has to be stored in the knowledge base.

Knowledge acquisition by interviewing soil experts proved to be difficult. The present generation experts have not yet been accustomed to express there knowledge in a way compatible with rule based knowledge engineering. This factor and the requirement for local knowledge led the authors (physical geographers) to select an area in Holland familiar to them. As the extend of the area was defined by the available SPOT scene (60km x 60km) a knowledge base could be set up which was not too large and still of practical use for the case study.

After generalizing the information from soil maps of the Netherlands (scales 1:50,000 and 1:250,000 (STIBOKA 1979; 1985)) to a scale compatible with the major soil classes of the region, a relation matrix (Table 1) was constructed relating ten soil classes to nine observable features separable into 39 attributes. The names of the soil classes follow Soil Taxonomy (SOIL SURVEY STAFF, 1975; De Bakker, 1979).

Input into the system is defined by a segmented SPOT image. For the present the extraction of attributes of image segments (natural "landunits") is performed by visual interpretation.
Attributes can be spectral (image tone) or spatial (parcel size, parcel shape, ...).

THE INTERFACE :

The user is provided with a SPOT image scale 1:100,000 with a segment overlay. The segmentation has been prepared by trained image interpretors and is based on landscape features.

The student is asked to input observed features / attributes for each landunit. The system guides the student through a hierachy of observations according to a backward chaining search strategy. On request the system will show the log of input and trace the decision making process, and it will provide explanation text , in reaction to WHY (do you ask this question) and EXPLAIN (how could I answer the question) .

HANDLING UNCERTAINTY

- WHAT CAUSES UNCERTAINTY IN CLASSIFICATION RESULTS ?

1- Uncertainty associated with inexact evidence e.g. caused by noisy measure ments or the inaccuracy of the human decision making process. The last one could be quantified by accumulating the hit/miss ration (confusion matrix) for a certain interpretor working in his field of expertise.

2- Uncertainty associated with rules. This means that the assertion :

IF <attributes x1 .. xn> THEN <class C1>

is not true in all cases.

DISCRIPTION OF THE PROTOTYPES

In this case study different approaches to rule based expert systems have been tried. Thse have been implemented in models written in (micro-)PROLOG and ENVISAGE.

BOOLEAN DECISION RULES

According to tabel 1, each soil class has a list of attributes. Applying logical AND and OR functions to match the given attributes with the list for each class would produce a classification. The evaluation of a Boolean expression with

AND/OR terms can be represented in an AND/OR tree and is equivalent to a box classifier in Pattern Recognition. The classification schema of figure 2 is equivalent to a set of production rules like :

IF "land-use is forest OR heath" AND "drainage type is no-superficial-drainage"
THEN "class is Haplortod"

These Boolean decision rules are eqeivalent to PROLOG Horn-Clauses (DE SARAM, 1985) :

```
((soil Haplorthod)
 (land-use forest OR heath)
 (drainage no-superficial-drainage))
```

Boolean rules suggest certainty, because the evaluation of the logic expressions produce only FALSE or TRUE.

MAXIMUM LIKELYHOOD

Instead of working with TRUE and FALSE in real life one has to do with likelyhoods. Furthermore decision rules have to include weight factors for the costs of false decisions and the benefit of good decisions. The maximum likelyhood (cost weighted) decision rule fullfils these requirements. It also allows the accumulation of evidence by using previous classification results to provide the prior probabilities in the BAYES rule :

$$P(Ci|x) * P(x) = P(x|Ci) * P(Ci)$$

where :

Ci class i , x = (x1 .. xn) observation tupple, attribute list.

For a given constant $P(x|Ci)$ and area dependant $P(Ci)$ the maximum likelyhood rule selects the class Cm with the maximum likelyhood . Storing the likelyhood vector $P(Ci')=P(Ci|x)$ in the database allows the accumulation of evidence when new components of the observation tupple (new attributes) become available.

- LIKELYHOODS IN A BOOLEAN ENVIRONMENT :

In GIS environments it is usual to preclassify ordinal data into nominal data by mapping the ordinal data into intervals. Mapping ordinal data into intervals followed by an AND function is equivalent to a box classifier decision rule. Overlapping boxes in feature space are equivalent to OR expressions connecting AND boxes. In figure 3 a three dimensional feature space partitioned into boxes is shown. The value of P(x|Ci) is stored for each soil class Ci. For a given P(Ci) the likelyhoods P(Ci|x) are also known for each box. Under the maximum likelyhood decision rule the classification result for each box is also defined. Consultation of the expert system is now equivalent to determine the box address from the observations (x1 .. xn) and read the corresponding class.

As the box classifier contains Boolean AND functions it can also be represented by PROLOG clauses. The difference with the Boolean approach is that in the learning fase frequences of co-occurance are used for the definition of the terms in the Bayes rule and the maximum likelyhood is explicitly stored as an attribute of each classified segment.

In this study PROLOG was used for the storage of the likelyhoods in a six dimensional attribute (feature) space :

(((x1..x6) 'soil-class' P(Ci|(x1..x6))))

where (x1..x6) is interpreted as a library label pointing to the content of a cell in six dimensional feature space. Redundancy in the featurespace is apparent from the fact that of the 6480 boxes in featurespace only 300 contained information. PROLOG has the nice property of compressing the sparsely populated likelyhood lookup-table into a list with 300 terms.

For a different area with different P(Ci) the classification lookup-table has to be recalculated from the generic relation :

(((x1..x6) 'soil-class' P((x1..x6)|Ci)))

and from the new P(Ci) according the the Bayes rule.

- FUZZY LOGIC ?

For an N-dimensional featurespace vector the rule for obtaining the certainty of fuzzy-AND is to take the minimum P(class|xi) over i. For fuzzy-Or the maximum over i is taken. As the minimum over all projections is less than the average over a box and the maximum over all projections is more fuzzy

logic is not supported by statistical reasonining.
Figure 4 illustrates the difference between statistical correct likelyhoods and fuzzy AND/OR rules for certainty estimation.

- LIKELYHOODS IN ENVISAGE

The ENVISAGE shell offers AND/OR trees, Fuzzy logic and a form of likelyhood calculatioh. The likelyhood calculations are implicitly based on the assumtion that all observations are statistically independant. This assumption is rather unrealistic and in practice the method is statistically weak. The statistical parameters are not derived from a co-occurance matrix but the user has to express a measure of belief (odds) in the answer he provides. After sevral sessions playing with the intuitive likelyhoods may produce acceptable results but unexpected results will still occur. Our measure of belief in the systems results remained relative low.

MODE OF INFERENCE :

Both the Boolean model and part of the ENVISAGE model use a mode of inference (reasoning) called backward chaining. This means that the inference process is started with the most likely hypothesis based on prior class probabilities. The most likely hypothesis will define the first querry to the user or to a database. The querry process will continue until either there is sufficient evidence for the hypothesis under consideration or there is insufficient evidence. In the last case the reasoning system will redirect its attention to a more likely hypothesis given the accumulated list of observations so far. The main advantage of backward chaining is that not all possible observations (attributes, features) have to be generated and examined for all classes.

CONCLUSIONS :

- The stimulus provided by applying knowledge engineering to the art of image interpretation was remarkable. It motivates a better contact between the staff specialised in photo (image) interpretation and the staff responsible for digital techniques in remote sensing. For future knowledge engineering projects it is desirable that application domain experts learn to program at the level of expert systems shells.

- Replacing basic tutorial sessions by computer assisted instruction is cost effective once the building of the knowledge bases has become a routine.
- These kind of exercises demonstrate a direct path from remotely sensed data to a GIS environment enhanced with certainty factors and explanation facilities. Users of GIS's have to be trained in thinking and do-ing in terms of likelyhoods and cost/benefit analysis.
- The knowledge concerning the relations between landscape elements and soils was less general than had been expected. An expert system can only be implemenmted for a limited domain. It is better to combine a number of low level expert systems in a hierarchical way than try to put a large and too complicated knowledge base in one large model.
- In any future operational system of knowledge engineering historical economical and social effects on observable features have to be taken into account. This means that any rule base for Earth resource applications will contains a limited body of general knowledge and a large body of local knowledge in the form of context maps and context rules.
- Models for e.i. soil genesis have to be included explicitly in the procedural part of the knowledge base, using e.g. the digital elevation data as a parameter file.

FUTURE RESEARCH

- In the near future the critical step of segmenting the image into landunits will be more automated.
- spatial features which now must be observed by the student interpretor will gradualy be extracted by pattern recognition routines. As these routines may require large computing resources, the knowledge engineering approach allows selective calling of subroutines based on the current state of the backward chaining inference process.
- next generation GIS must be developed supporting operational use of certainty factors (likelyhood vectors) and cost/benefit analysis applied to the decision making process.
- more features will be stored in a GIS, this will also reduce

the required input from the user.
- the interface between knowledge engineering systems KES and image processing and pattern recognition must be greatly improved. Present KES are too limited to "talking".

REFERENCES

Bakker,H.de, 1979; Major soils and soil regions in the Netherlands. Soil Survey Institute, Wageningen.
Mulder,N.J., 198 ; Classification and decision making, Photogrammetria, 40 95-116, Elsevier, Amsterdam.
Saram, H. de, 1985; Programming in micro-PROLOG, Ellis Northwood ltd.
Soil Survey Staff,1975; Soil Taxonomy, Soil Conservation Service, U.S. Dept of Agriculture.
STIBOKA, 1985; Bodem kaart van Nederland 1:250,000. STIBOKA, Wageningen.
SYSTEM DESIGNERS, 1985; ENVISAGE Reference Manual, Penbrooke.

features	landuse							rel.-elevation	
attributes / soil-classes	forest	heath	grass	arable	mixed-grass-arable	mixed-forest-arable	mixed-forest-grass	more-than-5-meters	less-than-5-meters
Haplorthod	+	+	-	-	-	-	-	+	-
Udipsamment	+	+	-	-	-	+	-	-	+
Plaggept	-	-	-	+	+	-	-		
Plaggeptic-Sphagnofibrist	-	-	-	+	+	-	-	-	+
Typic-Sphagnofibrist	+	+	-	-	-	-	+	-	+
Histic-Humaquept	-	+	+	-	+	-	-	-	+
Fluventic-Eutrochrept	-	-	-	+	+	-	-	-	+
Typic-Fluvaquent	-	-	+	-	-	-	-	-	+
Typic-Humaquept	-	-	+	-	+	-	-	-	+
Typic-Haplaquod	+	-	-	-	+	+	+	-	+

features	drainage-type				parcel-size			
attributes / soil-classes	river	brook	canals	no-superficial-drainage	larger-than-average	average	smaller-than-average	not-uniform
Haplorthod	-	-	-	+				
Udipsamment	-	-	-	+				
Plaggept	+	+	-	+	+	+	-	-
Plaggeptic-Sphagnofibrist	-	-	+	-	+	-	-	-
Typic-Sphagnofibrist	-	-	+	-	+	-	-	-
Histic-Humaquept	-	+	+	-				
Fluventic-Eutrochrept	+	-	-	-	+	-	-	-
Typic-Fluvaquent	-	-	+	-	-	+	+	-
Typic-Humaquept	-	+	-	-				
Typic-Haplaquod	-	+	-	-	-	+	+	+

features	parcel-shape					parcel-composition					
attributes / soil-classes	square	rectangular	elongated	polygonal	not-uniform	random	regular-blocks	irregular-blocks	parallel	circular	following-meanders
Haplorthod											
Udipsamment											
Plaggept	+	+	-	-	-	-	+	+	+	+	-
Plaggeptic-Sphagnofibrist	-	+	-	-	-	-	+	-	+	-	-
Typic-Sphagnofibrist	-	+	-	-	-	-	+	-	+	-	-
Histic-Humaquept	+	+	+	-	+	-	+	+	+	-	-
Fluventic-Eutrochrept	-	-	-	+	+	+	-	+	-	-	+
Typic-Fluvaquent	-	+	+	-	-	-	+	-	+	-	-
Typic-Humaquept	-	-	-	+	+	+	-	-	-	-	+
Typic-Haplaquod	+	+	-	+	+	+	+	+	-	-	-

features	tone-of-soil				village-type			situation-of-segment			
attributes / soil-classes	black	dark	normal	light	linear-settlement	other	no-village	part-of-an-elliptic-structure	situated-along-a-river	borders-a-forest-or-village	borders-a-brook
Haplorthod	-	-	+	+	-	+	+				
Udipsamment	-	-	+	+	-	+	+	+	-	-	-
Plaggept	-	+	-	-	-	+	+	-	-	+	-
Plaggeptic-Sphagnofibrist	+	+	+	-	+	-	+				
Typic-Sphagnofibrist	+	+	-	-	+	-	+				
Histic-Humaquept	-	+	+	-	+	-	+				
Fluventic-Eutrochrept	-	-	-	+	-	+	-	-	+	-	-
Typic-Fluvaquent					-	-	+	-	+	-	-
Typic-Humaquept								-	-	-	+
Typic-Haplaquod	-	-	+	-	-	+	-				

\+ = TRUE

\- = FALSE

empty cell = indifferent (OR-arc in AND/OR NETWORK)

1 Soil_class - attributes matrix for Overijssel.

Backward chaining mode of reasoning for 3 soil-classes with 3 features

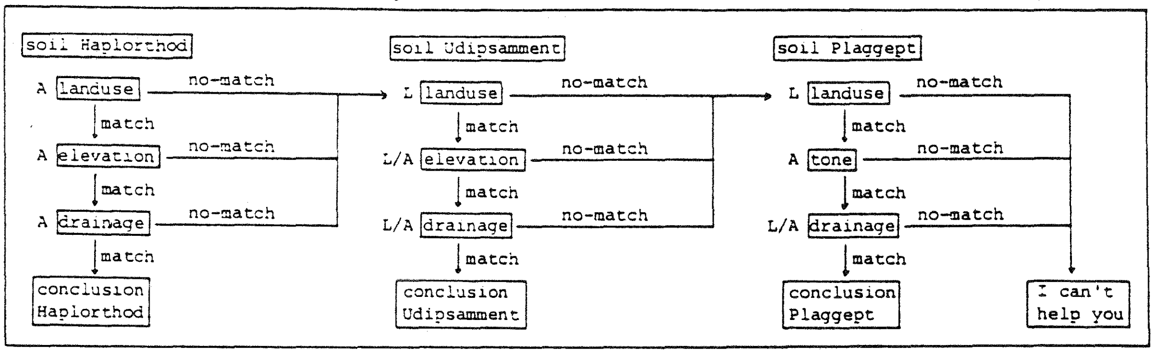

Fig. 2 Backward chaining

A: ask
L: look up
L/A: look up else ask

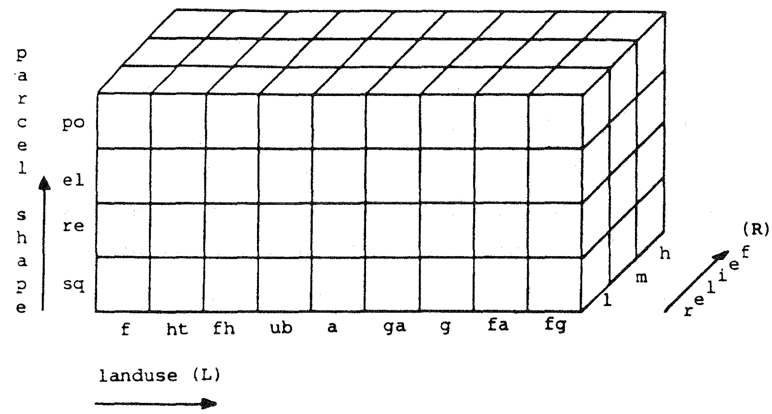

figure 3 The three-dimensional feature space.

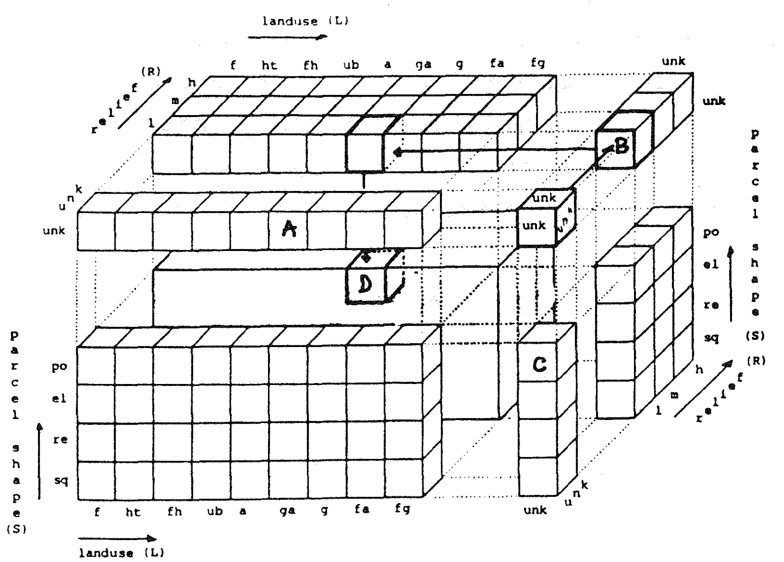

figure 4 Feature spaces in situations where one or more features are unknown.
The statistical model takes $P(Cl|x_1..x_n)$ from cell D
the fuzzy logic rule takes the minimum $P_{min}(C_i|x_j)$
of the boxes A, B and C.