OIL-WELL DATA INTERPRETATION USING EXPERT SYSTEM AND PATTERN RECOGNITION TECHNIQUE

Alain Bonnet
Ecole Nationale Superieure
des Telecommunications
46, Rue Barrault
75013 Paris, France

Claude Dahan
Schlumberger
12, Place des Etats-Unis
92541 Montrouge, France

ABSTRACT

This paper describes a program for interpretation of oil-well measurements. Two techniques have been merged for this purpose: the expert system technique (production rules) which lends itself to the symbolic reasoning aspect of the problem and pattern recognition technique (mainly the split and merge algorithm) to interpret curves measuring physical properties of rocks. Part of the knowledge of an expert in geology has been coded into 500 production rules together with hierarchy trees (such as petrography, paleontology) which allow the program an access to the rules containing assertions more general than the input data. There are two geological laws underlying the design of the program: (1) All sources of data are accounted for whenever they are available, (2) A description of a rock has no significance out of its context, which means that the identification of rocks must be guided by the knowledge of the environmental conditions. A session with LITHO is partially displayed.

INTRODUCTION

A crucial problem in petroleum exploration is the interpretation of measurements of certain physical properties of underground rocks such as density/ electrical resistivity, sound transmission, radioactivity. These measurements are represented by curves called logs which are characteristic of the various rocks penetrated by the drill.

In order to achieve their task of interpretation, geologists use various sorts of knowledge. They take into account qualitative data (e.g "cambrian" has a conotation of "very old" and therefore suggests "a very compacted rock") as well as quantitative data such as permeability, porosity, percentage of calcite which are complementary to the logs numerical values themselves.

Statistical approaches used so far to interpret geological data have drastic limitations. These limitations are due to the fact that some rocks (e.g. cherts and quartzites) have similar log responses and therefore can only be distinguished on the ground of their geological environment (geography, stratigraphy, paleontology).

LITHO is an ongoing project developed at Schlumberger (Clamart and Montrouge, France) for computer interpretation of various data recorded during and after the drilling phase: cores, cuttings, geographical data, seismic information, X-ray measurements, plugs, drilling parameters and logs data (curves reflecting the physical measurements mentioned above).

The output obtained by LITHO is a "litholog" which means (in petroleum industry jargon) a description of the rocks encountered in a well (vertical lithological column). For the time being, we are concerned with a "labelling" of the most plausible lithofacies (*) in terms of depth.

We will describe the different sources of data usually available and how they are actually used by the program. The fundamental hypothesis underlying the global project outlined here is that a high amount of geologic knowledge is indispensable (together with wire-line measurements knowledge) to solve the problem; furthermore overlapping information is the best guarantee against possible erroneous data.

(*) Lithofacies: the sum of petrophysical and petrographical aspects of rocks. The porosity, permeability, chemical and mineral composition of the rock, the arrangement of the different grains and particles (texture), the type of layering, presence of lenses, pebbles, laminations, holes, cracks, fissures (structure) define the lithofacies. An electrofacies is the image of a lithofacies in the space of the logs. A paleoenvironment is defined by all the climatic and geographical conditions that lead to a certain environmental context of deposition (e.g reefal, deltaic, arid).

RELATED WORK AND TECHNICAL DETAILS

LITHO is related to HASP/SIAP [5] which uses knowledge-based techniques for signal processing. Ancestors of LITHO as expert systems in geology are PROSPECTOR (4] and the DIPMETER ADVISOR [3]. PROSPECTOR was primarily concerned with hard-crystalline rocks whereas the emphasis in LITHO is on sedimentary rocks. Log interpretation has been already attempted with the DIPMETER ADVISOR but only one log was interpreted in detail. Here, all the logs are accounted for (sonic, density, resistivity, gamma-ray, etc.); information extracted can be overlapping or conflicting (sources of measurement errors are multiple). The strength of the approach of LITHO is the capacity of combining numerous sources of information (both symbolic and numeric) and accomodating possible contradictions.

Furthermore, most of the knowledge about geology is very general (for example geography independent).

LITHO is a new application of the EMYCIN framework after MYCIN[7], SACON {1] among others. These systems rely strongly on the context-tree mechanism; on the other hand, LITHO makes a more extensive use of the certainty factors combination functions. The latter aspect seems to be inherent in geology where most of the inferences drawn are weak and where contradictions between data are not uncommon.

LITHO is a complete interactive program using reasoning mechanisms acquired from human experts. It has the initiative of the dialogue through which it acquires the data characterizing the well which has been drilled. LITHO now runs under two versions: the DEC-20 version is written in INTERLISP and uses EMYCIN (8] [91; the VAX version uses ELITHOr a slightly simplified version of EMYCIN written in FRANZLISP. A FORTRAN (batch) program run on a VAX currently gives the final output (vertical geology description) on a VERSATEC.

SOURCES OF DATA

The primary goal of LITHO is to determine a set of most plausible lithofacies for each gross zone for example CHERTY LIMESTONE, TILLITE, FERROGINEODS SAND, THIN INTERBEDDING OF VARVES, EPSOMITE, etc.

The values of the goal parameter are inferred using data parameters asked for during the consultation (geologic observations, morphologies of curves, etc.) and using various intermediate steps like the main lithologic type, the paleoenvironment of deposition. The depth of the inference tree is about 4. In order to illustrate the combination of pattern recognition and expert systems techniques, this paper will concentrate on the two main sources of knowledge which are the external (geological) data and the shapes of the curves.

Morphology of logs

This is a primary source of data; there are two problems to be solved: first, the extraction of the important features of logs, second their interpretation within the reasoning mechanisms. This section will describe first the vocabulary that we have developed to describe the curves, then how these features intervene in the reasoning and finally the technique of pattern recognition used to extract these features.

Interesting features: The curves can be described in terms of whether there is indication of activity (flat or not), presence of plateaus (gamma-ray, resistivity, neutron, etc.), ramps (gamma-ray, etc.), beds (microresistivity, gamma-ray, etc.), layers (HDT, microresistivity), strings (HDT only). From these data, the program infers the most plausible environments, main lithologic types and sometimes directly plausible lithofacies. As the pattern recognition program can miss micro-events which are difficult to recognise, the user has the possibility to complete the feature extraction by interpreting visually the logs. For this reason, s/he has a handbook providing some help for the recognition of the various shapes. Hesitations do not block the deduction process since the user is permitted to introduce credibility factors weakening the inferences to be made. Some of the shapes taken into account are listed below. (The observer is always looking upwards):

Plateaus: Constant log response over an interval longer than 25 meters. Megaramps: progressive increasing or decreasing response longer than 25 meters. Massive beds: Constant log answer over an interval comprised between 5 and 25 meters.

An example of a (nearly) pure rule using the morphology of the logs is shown below:

RULE003:

If: 1) There is one or several plateaus on the FDC curve, and
2) There is one or several plateaus on the CNL curve, and
3) There is one or several plateaus on the BHC curve, and
4) The global porosity of the zone is less than 10 percent

Then: There is suggestive evidence (.5) that the geological formation of the zone is globally compacted.

Morphological analysis of logs:

The major problems for an automatic pattern recognition system are of two kinds :

-Despite the apparent simplicity of the ideal representation of a given feature, the variation tolerated around the ideal curve is very large.

-A given structure may be split into many substructures (i.e: a megaramp may include medium ramps, medium beds, etc.)

This last situation is natural for a geologist who adapts his scale of vision to the details relevant for his appreciation of the geological context.

For these reasons, the following approach has been chosen; it consists of 2 steps:

-A log polygonal approximation

-A syntactic parsing of the approximating polygone Logs approximation:

The piecewise linear approximation of a log is interesting for the two following reasons:

-The approximation can be matched with the scale of details which seems relevant (different scales are easily considered)

-Each curve part is easily coded (two numbers for a segment)

We have restricted ourselves to piecewise approximation with no continuity requirements at the junction between two segments. It turns out that this limitation is actually beneficial. This is due to the fact that it is of primary importance to locate ruptures in the signal ("discontinuities"), which will be emphasized if no continuity constraint is taken into account.

Such approximations are well suited to our problem because the degree of detail obtained in the approximating polygone is easily controlled. They are reached by using a split-and merge algorithm (6J.

Parsing of the logs:

The curve decomposition into elementary segments is suited to a linguistic approach for the pattern recognition problem. In this type of processing, the waveshapes to be identified are viewed as a set of vectors concatenation which are analogous to the words of a language generated by a given grammar. The method is the following:

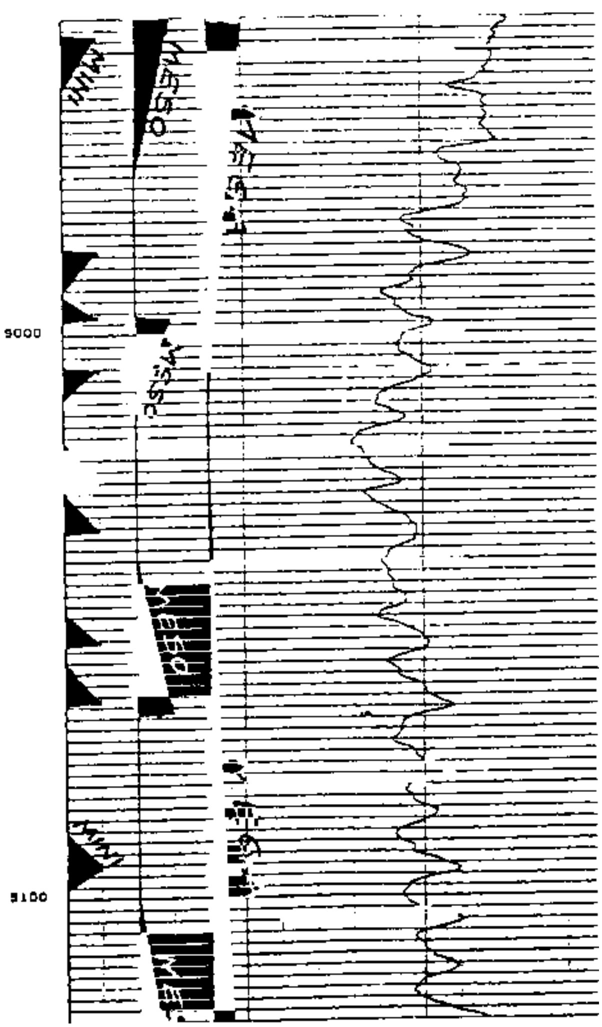

Fig 1. Feature extraction on the density log

-Starting from the words corresponding to the desired curves shapes lead to define a grammar which will generate exactly those words

-Constructing the corresponding parser, which will decide, examining a vector arrangement, whether it is acceptable by the grammar.

In this formalism, for instance, a decreasing ramp is represented by "the concatenation .of negative slope segments, the first slope segment being under a given threshold, and the jump between two successive segments (produced by the approximation) being under another threshold"; the parser that matches the feature against the grammar is implemented as a finite automaton.

An example of performances of the overall procedure -polygonal approximation and syntactic parsing - is illustrated in figure

External data

The term "external data" will refer to all information which does not come from wire-line measurements. Some are not always available for example cutting descriptions or paleontological data. Others are always known (or can be inferred) such as the location of the well bearing very useful information about the global lithologic type usually encountered in a given area (e.g platform-carbonates in the Middle-East). Usually, age of formations (stratigraphical knowledge) is partly known from the geographical information.

External knowledge is of 8 types outlined below. It either consists of taxonomies enabling accers to general rules from specific data or of inferences triggered as information is entered (if the age of the formation is "cambrian" then the rocks are very old (older than 185 millions of years) suggesting a high probability of very compacted rocks, i.e no gypsum, etc.). The 8 kinds are:

Geography (geologic provinces, basins, fields)
Tectonic activities (folds, faulting)
Stratigraphy (geological eras, epochs)
Paleontology (fossiles)
Mineralogy (calcite, clay, quartz)
Petrography (limestone, shale, sand)
Sedimentology (reef, tidal channel)
Petrophysics (porosity, permeability)

An example of rule combining geography and stratigraphy information is shown here:

RULE295:

If: 1) The stratigraphical period is cretaceous,
and 2) The geological province is one of:
iranian-platform, zagros-fold-belt
Then:
1) There is evidence that the paleoenvironment of deposition of the zone is fluvial (.6), lacustrine (-.6), deltaic (-.6), glacial (-.8), arid (-.6), lagoonal (-.3), pelagic (.4), neritic (.4), metamorphic (-.6), intrusive (-.8), extrusive (-.3), karstic (-.3), coastal (-.3) or reefal (.2)

2) There is evidence that the main lithologic type of the zone is detrital (-.3), biological (.3), evaporitic (.3) or plutonic (-.5)

Zoning and labeling of the well:

This last step is essentially an allocation procedure in the log values space. This space, for the circumstance, is filled with the clusters of points. A point is a collection of log values which make up the data base already used in the rough lithological survey at the beginning of the processing.

Each cluster is representative of a lithofacies, and the symbolic reasoning part of the program provided us with the a priori plausibility of the presence of this lithofacies.

The allocation procedure used in this last phase is once more of the nearest-neighbour type, the (Euclidean) distances being weighted by the previous plausibilities.

A final output of the whole chain is the result of this last program, in the form of a lithological column as shown in Figure 2. This comparison shows two things:

First the drawings made by the geologist and by LITHO are extremely similar; their agreement is actually beyond our initial hopes.

Second LITHO gives a terse description of the lithology which is usually a good summary of the annotations entered by the geologist.

How does Litho handle spurious date? Spurious data may have diverse origins:

Logs are never error-free. This is due to the conditions of drilling which are never optimal.

There may be errors on the cuttings (a piece of ground may have fallen down during the operation of drilling thus causing its attribution to a wrong place).

Litho copes with these problems oy using the following features:

There is never a full certainty attached to the conclusions drawn. Therefore if data responsible for a given conclusion happens to be wrong, other sources of information might help the system to still reach a reasonable conclusion.

It makes a heavy use of overlapping knowledge. A strong conclusion often results from the combination of 5 to 10 weak conclusions.

EXAMPLE OF A SESSION WITH LITHO

The following transcript illustrates LITHO's interactions with a user and the conclusions of the program faced with a real case. The user inputs appear after the double asterisks. Comments have been inserted between brackets.

```
----------WELL-106----------
1) Name of well-106
** DIDI-7
2) Where are you?
** IRAQ-NORTH
The geological province is ZAGROS-FOLD-BELT
The continent is MIDDLE-EAST
 [Litho uses geography knowledge in order
 to infer the geological province and
 continent from the geological basin.]
3) Are there any zones of DIDI-7 that
   you wish to study?
 [Litho usually works on "gross" zones
 defined by the user.]
** YES
   ----------ZONE-1----------
4) What is the geological age of DIDI-7?
** UNKNOWN
5) Please enter the paleontology words
   that you may know for ZONE-1
** FORAMINIFERA OSTRACUDS
= OSTRACODS   [spelling corrector]
The paleontological classes of ZONE-1 are:
FORAMINIFERA ARTHROPODS
The paleontological order of ZONE-1 is:
OSTRACODS
 [Use of taxonomies (class, order, family,
  genus) to make inferences.

14) Please enter the characteristic features
    of the gamma-ray log in ZONE-1
plateaus: NO
massive beds: NO
megaramps: ABSENT
mesoramps: ABSENT
medium beds: NO
medium ramps: RARE
 [The user looks at the outputs of the
 pattern recognition program to answer]

23) What is the tectonic structure of ZONE-1?
** WHY
This will aid in determining the
paleoenvironment of deposition of ZONE-1.
It has already been established that
[1.1] the main lithologic type of
ZONE-1 is biological, therefore if
[1.2] fracture is one of the tectonic
structure of ZONE-1, then
There is weakly suggestive evidence (.3)
that the paleoenvironment of deposition of
ZONE-1 is reefal (also: there is weekly
suggestive evidence (.3) that the
geological formation is globally compacted)
[The user asks WHY (why do you want to know
the tectonic structure?); Litho displays the
current rule being applied together
with the partial results obtained thus far.]

Here are my conclusions:
The paleoenvironment of deposition is
COASTAL or LAGOONAL
The main lithologic type is BIOLOGICAL
The most plausible lithofacies are:
WELL CEMENTED DOLOMITIC LIMESTONE
WELL CEMENTED DOLOMITE
MODERATELY CEMENTED DOLOMITIC LIMESTONE
ANHYDRITIC DOLOMITE
ARGILLACEOUS LIMESTONE
[LITHO gives a few intermediate conclusions
as well as the most plausible rocks in the
studied area]
```

CONCLUDING REMARKS

Assessment of Litho has been satisfactory on about 10 zones of different wells e.g. Bul-Hanine (Qatar), Didi (Syria), Prinos

(Greece), North-West Dome (Qatar), Rehden (Germany). Indeed, we believe that the power of the whole project relies on the combination of the two techniques mentioned above.

There are two geological laws underlying the design of the program LITHO: all sources of data are accounted for whenever they are available. Since a rock has no significance out of its context, the identification of a lithofacies must be guided by the knowledge of the environmental conditions. The latter is probably true in other domains such as medicine where the interpretation of, say electrocardiograms must be performed using contextual information about the patient.

Further research and improvement of the program might include:

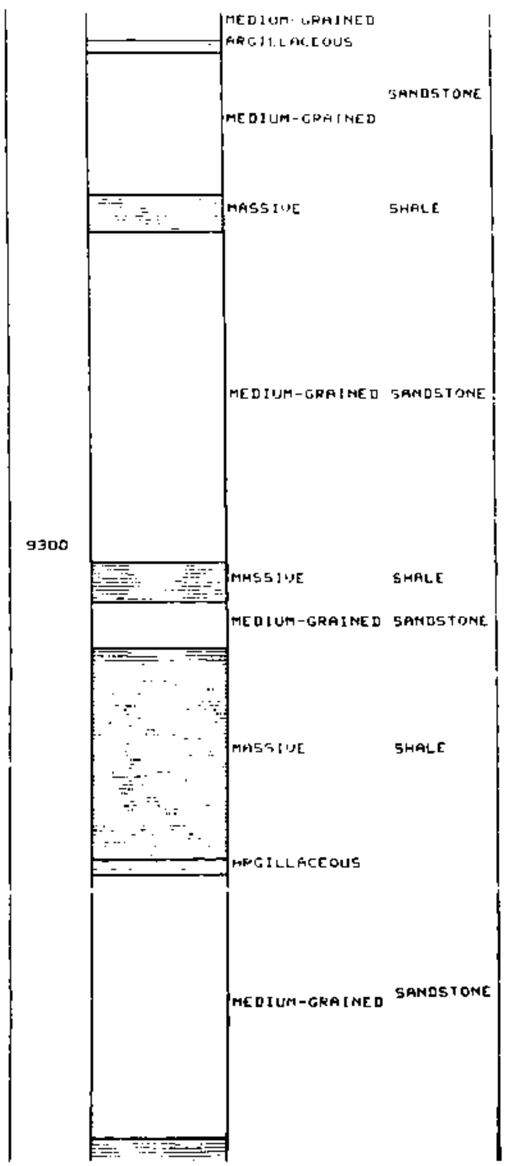

Fig 2. Final output of LITHO

* Developing an elaborate model of regional data for oil exploration.

* A possible extension of the program will include the determination of the depositional genesis of rocks (e.g "mouth bar sandstone" instead of "sand"). In other words we plan to determine the FACIES themselves.

We hope that experience with the program will give us a good insight into how geologists solve problems of data interpretation. In fact we envisage a further use of the program by students willing to learn log interpretation since most of the inference mechanisms can be appropriately displayed upon the user's request.

ACKNOWLEDGMENTS

We are indebted to Frederic Bastenaire, Jim Bennett, Bruce Buchanan, Ed Feigenbaum, Jean Gabriel Ganascia, Jacques Barry (our expert in geology), Yves Kodratoff, Penny Nii and Carli Scott with whom we had very helpful discussions when designing or implementing the program.

REFERENCES

1. Bennett J.S et al., SACON, a knowledge-based consultant for structural analysis, Technical report HPP-78-23, Computer Science Department, Stanford University, (September 1978).

2. Davis R., King J., An overview of production systems, in E. W. Elcock and D. Michie (eds.), Machine Intelligence 8, John Wiley and sons, New-York, (1977).

3. Davis R., Austin H., Carlbom I., Frawley B., Pruchnik P., Sneiderman R., Gilreath J.A., The dipmeter advisor: Interpretation of geologic signals, 7th International Joint Conference on Artificial Intelligence, Vancouver, (August 1981).

4. Duda R., Gaschnig J., Bart P., Model design in the Prospector consultant system for mineral exploration, in Expert systems in the microelectronic age, D. Michie (ed.), University of Edinburgh Press, (1979).

5. Nii P. H., Anton J. J., Feigenbaum E. A., Rockmore A. J., Signal to symbol transformation: HASP/SIAP case study, The AI Magazine, 23, (Spring 1982).

6. Pavlidis T., Structural pattern recognition, Springer-Verlag (ed.), (1980).

7. Shortliffe E.A., Computer-based medical consultations: MYCIN, Elsevier computer science library, (1976).

8. VanMelle W., A domain-independent system that aids in constructing knowledge-based consultation programs, Stanford University Heuristic Programming Project memo 80-22, (June 1980).

9. VanMelle W., Scott A.C., Bennett J.S., Peairs M.A., the EMYCIN manual, Computer Science Department Report HPP-81-16, Stanford University, (November 1981).