

PRELIMINARY PERFORMANCE ANALYSIS OF THE PROSPECTOR
CONSULTANT SYSTEM FOR MINERAL EXPLORATION

John Gaschnig
Artificial Intelligence Center
SRI International
Menlo Park, CA 94025

Abstract

To demonstrate that the performance of an expert knowledge-based system is comparable to that of the experts it emulates, it is useful to subject the system to an appropriate objective evaluation. The Prospector consultant system is intended to aid a geologist in evaluating the mineral potential of an exploration site. Here we report the results of a preliminary performance analysis of three Prospector ore deposit models. Using data from known deposits as test cases, we compare the system's performance in detail with analogous target values supplied by the model designer based on the same input data. These calibration results measure how well a model embodies the model designer's intentions, and identify particular sections of a model that would benefit from revision. We discuss limitations of the present experiments and future work. Put briefly, we report how work-a-day performance analysis instruments can accelerate the model design and refinement process in expert systems.

1 Testing Prospector Models

Performance analysis of expert systems can serve several useful purposes: (1) to demonstrate objectively and convincingly that the competence of a system is (or is not) comparable to that of the experts it emulates; (2) by exposing the inner workings of a complex system and knowledge base, to aid the ongoing efforts of the system implementors to achieve a level of performance that merits a formal evaluation by outside experts. Performance analysis techniques can be applied to systems of diverse domains of expertise, such as medical diagnosis (e.g., [10]), speech understanding (e.g., [6]), and game playing (e.g., [4]). In this paper we report another instance of the performance analysis of an expert system -- the Prospector consultant system for mineral exploration. The present work emphasizes the work-a-day use of performance analysis as an aid to system implementors in refining a knowledge base. The performance data provide an objective, detailed, quantitative measure of the current performance of a model, and pinpoint those sections of the model that are not performing exactly as intended, thereby establishing priorities for future revisions. Hence besides reporting specific experimental results for the Prospector system, the present work provides an example of an experimental methodology that may be adapted to (or improved upon in adapting it to) other expert systems.

The Prospector system is intended to emulate the reasoning process of an experienced exploration geologist in assessing a given prospect site for its likelihood of containing an ore body of the

type represented by the model he or she designed [2] [3]. Like other rule-based expert consultation systems (e.g., [8], [5], [7]), Prospector engages the user in a dialogue, accepting volunteered observations and prompting the user with questions about relevant factors not volunteered. Here we report the results of preliminary performance measurement experiments for a Prospector ore deposit model concerning porphyry copper deposits (denoted PCDA).

The basis for a Prospector computation is a set of ore deposit models formulated by expert economic geologists. We perform a separate evaluation for each model. In this paper "Prospector" always means the Prospector system executing some particular model.

2 Experimental Methodology

We wish to compare Prospector to human geologists in terms of the accuracy with which the presence or absence of a particular type deposit at a particular location can be predicted, but there exist no standard quantitative measures of human performance. Lacking an absolute scale of comparison, we employ instead a relative scale; our principal objective here is to measure how closely Prospector's conclusions agree with those of the model designer.

To compare Prospector's conclusions with those of its model designer, we ask the model designer to express his or her conclusions on the same -5 to 5 certainty scale used by Prospector. This scale measures degree of subjective certainty: the value 5 indicates absolute certainty that a

geological characteristic is present or that a final or intermediate model conclusion is true; -5 indicates absolute certainty that a characteristic is not present or that a conclusion is false; zero indicates no information; values between 0 and 5 and between 0 and -5 denote degrees of uncertainty. Values on this subjective certainty scale map (in a piece-wise linear, one-to-one onto fashion) to probability values, which are used exclusively for computation purposes. (Prospector's reasoning about uncertainties is based on a modified Bayesian probability updating scheme [1]; see [9] for a description of the subjective certainty scale used in MYCIN.) Since Prospector models have a hierarchical structure, we wish to compare both Prospector's overall conclusion and its major subconclusions with those of the model designer.

The test sites chosen here are all exemplars of the PCDA model. The experimental procedure is as follows: (1) for each test site, the model designer completes a questionnaire listing all the questions (about geological characteristics of the test site) asked by the model; (2) for each such input data set, the model designer assigns to each of several major sections of the model a target certainty value, expressing his judgment as to the extent to which the conclusion represented by that section is established for that test site; (3) Prospector is run once for each input data set, and the final certainty values assigned to each node in the model are recorded; (4) these raw performance and target values are analyzed in ways described subsequently.

Additional details about the experiment methodology are given in [2]. The fact that all the present test cases are exemplars of their respective models, and the fact that the model designer supplied the input data concerning the test cases, are limitations; the present tests are more necessary than sufficient conditions for good performance. Despite these limitations, the preliminary results reported in subsequent sections have proved useful in the ongoing model refinement process.

3 Performance Analysis of the PCDA Model

We now evaluate one version of a model for a class of porphyry copper deposits (PCDA) designed by Prof. Marco Einaudi of Stanford University. Input data was available for three test cases, namely the known deposits called Yerington (Nevada), Bingham (Utah), and Kalamazoo (Arizona). Prospector scored 4.769, 4.721, and 4.756, respectively for these three sites, indicating acceptable performance of the PCDA model (since these three sites are considered exemplars of the PCDA model).

The target certainty values assigned by Prof. Einaudi for calibration purposes are given either in the form of a single number (on a -5 to 5 scale), or as two numbers establishing an upper and lower bound on a certainty interval. The estimates are listed in Table 1 on the left, with the scores as determined by execution of Prospector recorded on the right. For brevity, Table 1 lists data only for the Yerington deposit. The symbols PCDA, FPTS, FRE, etc., in Table 1 denote nodes in the PCDA model hierarchy corresponding to major sections and subsections of the model (indicated by indentation).

Table 1. Prospector Scores for Several Levels of the PCDA Model -- Yerington Deposit

Name of Model Node	Einaudi's Estimate	Prospector Score
PCDA	4.5 to 5.0	4.769
FPTS	4.5 to 5.0	4.528
FRE	4.5	4.540
FPCDAIS	4.5 to 5.0	4.787
FCDS	5	4.524
FIS	5	4.744
FAMR	4.5 to 5.0	4.225

In most cases shown in Table 1 Prospector agrees very closely with Prof. Einaudi's estimate. Conclusions about these data can be expressed quantitatively by first identifying the values in Table 1 with a concise notation, then defining a simple formula for the relative error of Prospector in predicting Prof. Einaudi's estimates. We let $C(X, Y, Z)$ denote the certainty score given to model node Z by agent X for site Y , where X denotes Prospector or Einaudi. For example, $C(\text{Prospector}, \text{Yerington}, \text{FPCDAIS}) = 4.787$. When Einaudi gave an interval of certainty values instead of a single value, we use the midpoint of the interval as the value of C . Then an error measure is given by

$$E(Y, Z) = \frac{C(\text{Einaudi}, Y, Z) - C(\text{Prospector}, Y, Z)}{C(\text{Einaudi}, Y, Z)}$$

For example, $E(\text{Yerington}, \text{FPCDAIS}) = (4.75 - 4.787) / 4.75 = -.008$, meaning that Prospector's prediction is accurate to within 0.8% in this case. Supplementing Table 1 with analogous data for the Bingham and Kalamazoo deposits, we can compute the value of E for $3 \times 7 = 21$ different instances. For 5 of the 21 data points Prospector predicted Einaudi's estimate to within 1%, while 15 of the 21 data points show agreement to within 10%. The grand average over the 21 data points is 10.3%. For convenience, we list these 21 values of E in Table 2, expressed as percentages. In Table 2, "Average" denotes the average of absolute values.

Table 2. Relative Error (E) of Prospector Scores as predictors of Einaudi's Estimates (derived from data in Table 1 and analogous data)

	Yer.	Bingham	Kal.	Average
PCDA	-3%	-4.9%	-11.9%	5.7%
FPTS	4.7	-18.6	-4.7	9.3
FRE	-9	-13.6	49.0	21.2
FPCDAIS	-8	.4	-9	.7
FCDS	9.5	51.9	5.6	22.3
FIS	5.1	5.1	5.1	5.1
FAMR	11.1	5.6	5.6	7.6
Average:	4.1	14.3	11.8	10.3

Note that the average data in the rightmost column suggest the FRE and FCDS sections as candidates for model revisions. More extensive data and interpretation are found in [2].

4 Conclusions and Future Work

We have applied a simple experimental methodology for measuring quantitatively the performance of Prospector ore deposit models in some detail. The results indicate that the PCDA model reflects fairly faithfully the judgments of its designer. Although covering a limited number of cases that include only exemplars of the model, these performance analysis results prove useful in identifying particular sections of a model that would most benefit from "fine-tuning", hence establishing precise priorities for model revisions. Directions for future work include analogous experiments on additional cases, sensitivity analysis of input data and rule strength values, and incorporation of performance analysis into Prospector's explanation system.

ACKNOWLEDGEMENTS

This work was supported in part by the Office of Resource Analysis of the U.S. Geological Survey under Contract 14-08-0001-15985, and in part by the RANN Division of the National Science Foundation under Grant AER77-04499. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author, and do not necessarily reflect the views of either the U.S. Geological Survey or the National Science Foundation.

REFERENCES

1. Duda, R. O., P. E. Hart, and N. J. Nilsson, "Subjective Bayesian Methods for Rule-Based Inference Systems," National

Computer Conference 1976 (AFIPS Conference Proceedings, Vol. 45), pp. 1075-1082, 1976.

2. Duda, R. O., P. E. Hart, P. Barrett, J. Gaschnig, K. Konolige, R. Reboh, and J. Slocum, "Development of the Prospector Consultation System for Mineral Exploration," Final Report, SRI Projects 5821 and 6415, Artificial Intelligence Center, SRI International, Menlo Park, California (October 1978) (1978a).
3. Duda, R. O., P. E. Hart, N. J. Nilsson, and G. L. Sutherland, "Semantic Network Representations in Rule-Based Inference Systems," Technical Note 136, Artificial Intelligence Center, SRI International, Menlo Park, California (March 1977). Appears also in Pattern-Directed Inference Systems, D. A. Waterman and F. Hayes-Roth (eds.) (Academic Press, New York, 1978) (1978b).
4. Gillogly, J. J., "Performance Analysis of the Technology Chess Program," Report CMU-CS-78-109, Dept. of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania (March 1978).
5. Martin, N., P. Friedland, J. King, and M. Stefik, "Knowledge Base Management for Experiment Planning in Molecular Genetics", Proc. Fifth IJCAI, Cambridge, Mass., August 1977, pp. 882-887.
6. Paxton, W. H., "A Framework for Speech Understanding," Technical Note 142, Artificial Intelligence Center, SRI International, Menlo Park, California (June 1977).
7. Pople, H. E., "The Formation of Composite Hypotheses in Diagnostic Problem Solving: An Exercise in Synthetic Reasoning," Proc. Fifth IJCAI, Cambridge, Mass., August 1977, pp. 1030-1037.
8. Shortliffe, E. H., Computer-Based Medical Consultations: MYCIN, Elsevier, New York (1976).
9. Shortliffe, E. H., and B. G. Buchanan, "A Model of Inexact Reasoning in Medicine," Math. Biosci., Vol. 23 (1975), pp. 351-379.
10. Yu, V. L., et al., "Evaluating the Performance of a Computer-Based Consultant," Heuristic Programming Project Memo HPP-78-17, Dept. of Computer Science, Stanford Univ., September 1978.