Introduction to Prospector

Tao Zhang
(zt@wpi.edu)

An ES in the Geology Domain

- Prospect Evaluation
- Regional Resource Evaluation
- Drilling-site Selection
- Training
Agenda

- System Overview
- Inference Network
- Modeling
- Semantic Network
- Test Results

Prospector Architecture: Overview

- Inference Engine
- Knowledge Base
- Active Memory
- Explanation Facility
- User Interface
Key System Data

- Developed during 1976-1981
- Key figures
  - Richard Duda, John Gaschnig, Peter Hart, Rene Reboh, Nils Nilsson
- Implemented with INTERLISP
- Run on DEC PDP-10 computer
- Total 300 pages of source code
- Consumed about 165 K memory (?)
- Involves roughly 10 man-years of effort

Mode of Operation

- Interactive consultation
  - Questioning
  - Explanations
  - Respond to user commands
- Batch processing
  - For testing purpose
  - Or, for consulting large region
- Compiled Execution
  - Runs 4 orders of magnitude faster
Vocabulary

- **Inference network**
  A generic method for representing judgmental knowledge; A simple language that an expert can use to specify both the knowledge and how that knowledge should be used.

- **Model**
  A body of knowledge about a particular domain of expertise encoded into the system which the system can act.

- **Semantic Network**
  A network of nodes linked together by directed arcs to represent relevant knowledge like taxonomic relations among objects in the domain.

Inference Engine: Advantages

- Same knowledge be used more than 1 purpose
- Allow a large system be developed incrementally.
- Applied to similar problem domains by replacing knowledge base.
Certainty and Probability

\[ C(H \mid E) = \begin{cases} 
5 \frac{P(H \mid E) - P(H)}{1 - P(H)} & \text{if } P(H \mid E) \geq P(H) \\
5 \frac{P(H \mid E) - P(H)}{P(H)} & \text{if } P(H \mid E) < P(H) 
\end{cases} \]

P(H) is the prior probability of any hypothesis in the absence of evidence.
P(H|E) is the posterior probability with the observation of a piece of evident E.
C(H|E) measures certainty value.

One-to-One Relation C<->P

\[ C(H|E) \]

\[ P(H|E) \]

CS538 Expert System
Interpretations of C(H|E)

-5 = certainly false  
-4 = very probably false  
-3 = probably false  
-2 = unlikely  
-1 = somewhat unlikely  
0 = no opinion  
1 = somewhat likely  
2 = likely  
3 = probably true  
4 = very probably true  
5 = certainly true

Problems to Estimate Posterior Probability with Evidence gathered

- The available evidence is generally incomplete and uncertain.
- The probabilistic relations link the hypotheses and relevant evidence are both unknown and complex.
Solution: Hierarchy Structuring

- The human expert will usually identify a small number of major considerations that more or less independently influence the decision.

- The determination of the state of these major factors is done through the same kind of breakdown into major sub-factors, leading to a hierarchical decomposition of the decision procedure.

What Inference Networks Do

- Provide a simple way to specify what the factors are and which affect which other.

- Provide a set of standard ways of computing the probability of a given factor from the probability of the factors that influence it.
Categories of All assertions

- Top-level hypotheses
- Intermediate factors
- Evidential statements.

IN Topology(1) – Tree

If only one path from any evidence node to any top level hypothesis, the network has a tree structure.
Multiple paths are not unusual. In this case the IN is a genuine graph.

“Inference Networks are Acyclic Graph”

To prevent “circular reasoning”, the presence of loops is forbidden.
IN Topology(4) – Undesirable Graph

Generally speaking whenever a node has more than 4 or 5 antecedents, it is desirable to create new intermediate factors that separate the interactions of these antecedents.

Relations between assertions

- Logical Relations
- Plausible Relations
- Contextual Relations
Combining Evidence:
Logical Combinations

- Conjunction
  \( A = A_1 \) and \( A_2 \) \ldots and \( A_k \)
  \[
P(A | E') = \min_i \{P(A_i | E')\}
  \]

- Disjunction
  \( A = A_1 \) or \( A_2 \) \ldots or \( A_k \)
  \[
P(A | E') = \max_i \{P(A_i | E')\}
  \]

Combining Evidence:
Weighted Combinations

- Prior Odds on \( A \)
  \[
  O(A) = \frac{P(A)}{1 - P(A)}
  \]

- Likelihood Ratio (LR), “Sufficiency Measure”
  \[
  \lambda_i = \frac{P(A_i | A)}{P(A_i | \overline{A})} \quad ("LS")
  \]

- Bayes’ rule states that:
  \[
  \log O(A | A_1, A_2, \ldots, A_k) = \log O(A) + \sum_{i=1}^{k} \log \lambda_i
  \]
Weighted Combinations (con’d)

- Bayes’ Rule assume $A_i$ is known true.
- If we only have $P(A_i|E')$ that $A_i$ is true, effective LR determined by 3 fixed points:

$$\lambda_i = \begin{cases} 
\lambda_i & \text{if } P(A_i|E') = 1 \\
1 & \text{if } P(A_i|E') = P(A_i) \\
\overline{\lambda_i} & \text{if } P(A_i|E') = 0 
\end{cases}$$

$\overline{\lambda_i}$ is the LR when $A_i$ is known false, “Necessity Measure”

$$\overline{\lambda_i} = \frac{P(A_i|A)}{P(A_i|\overline{A})} \ ("LN")$$

Contexts and Subgoals

- Designate any proposition $C$ as a *context*.
- Context arc $(A \rightarrow C)$ blocks the upward propagation of any info about $A$ if context hasn’t been established.
- If a conclusion depends on $A$, Inference Network will set up the subgoal of first establishing context $C$.
- Context mechanism goes beyond factual knowledge representation to *control*.
Model Revisited

- “A body of knowledge about a particular domain of expertise encoded into the system which the system can act.”
- Prospector consists of a number of such specially encoded models of certain classes of ore deposits.
- Intended to represent most authoritative and up-to-date info available about each deposit class.
Models in Prospector

- Performance of Prospector depends on
  - Number of models
  - Type of deposits modeled
  - Quality & completeness of each model
- By 1983, 23 models has been constructed
  - Consisting of 1800 nodes
  - 1370 rules

Models in Prospector (cont’d)

- Each model is encoded as a separate data structure independent of Prospector sys.
- The Prospector system should not be confused with its models.
- Prospector is a general mechanism for delivery relevant expert info to a user who can supply it with data about a prospect.
Model Development Process

- A. Initial Preparation
- B. Initial Design
- C. Installation and debugging of the model
- D. Performance Evaluation and Model Revision

Form of Knowledge Representation

- Taxonomies and Semantic Networks
  Basic concepts - rock types, minerals, ages, etc. are organized as a hierarchical tree structures with simple relationships (e.g. subset/superset)
  - Then be combined using domain specific relations to form more complex statements
    - Represented by partitioned semantic networks.
Semantic Networks Enable the System to

- Recognize & exploit general taxonomic relations
- Interconnect different models automatically
- Connect user supplied information to the models

Comparing with the Expert
Average difference is 0.69, or 6.9% of the –5 to 5 scale.

<table>
<thead>
<tr>
<th>Test region</th>
<th>Target value</th>
<th>Prospector score</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Hills</td>
<td>3.50</td>
<td>4.33</td>
<td>-0.83</td>
</tr>
<tr>
<td>Crooks Gap</td>
<td>4.70</td>
<td>4.26</td>
<td>0.44</td>
</tr>
<tr>
<td>Gas Hills</td>
<td>4.90</td>
<td>4.37</td>
<td>0.53</td>
</tr>
<tr>
<td>Shirley Basin</td>
<td>4.95</td>
<td>4.13</td>
<td>0.82</td>
</tr>
<tr>
<td>Ambrosia Lake</td>
<td>5.00</td>
<td>4.39</td>
<td>0.61</td>
</tr>
<tr>
<td>Southern Powder River</td>
<td>4.40</td>
<td>4.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Fox Hills</td>
<td>1.50</td>
<td>2.17</td>
<td>-0.67</td>
</tr>
<tr>
<td>Oil Mountain</td>
<td>1.70</td>
<td>3.32</td>
<td>-1.62</td>
</tr>
</tbody>
</table>
Conclusions

Inference networks effectively provide a formal language for the Expert System tasks and decision making