Introduction to Prospector

Tao Zhang
(zt@wpi.edu)

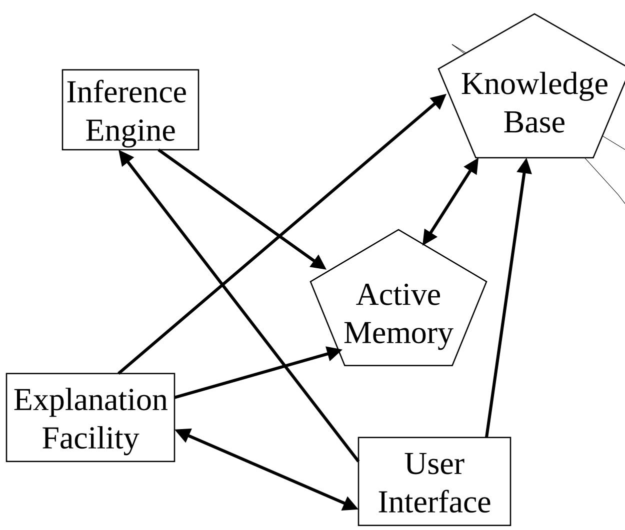
An ES in the Geology Domain

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

Agenda

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

Prospector Architecture: Overview



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

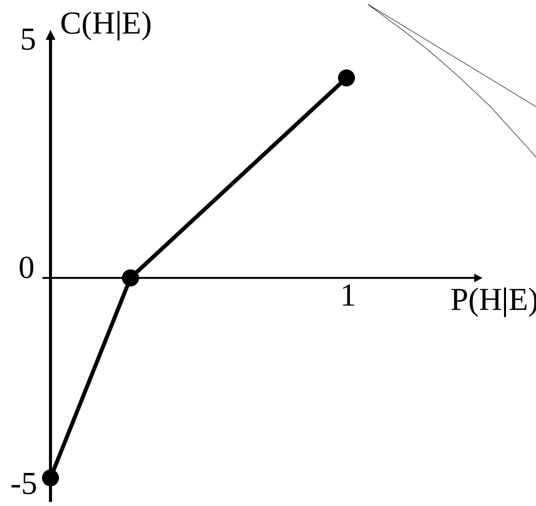
$$C(H|E)=\begin{cases}5\dfrac{P(H|E)-P(H)}{1-P(H)} & if\ P(H|E)\ge P(H)\\ 5\dfrac{P(H|E)-P(H)}{P(H)} & if\ P(H|E)< P(H)\end{cases}$$

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



Interpretations of C(H|E)

-5=certainly false	5=certainly true
-4=very probably false	4=very probably true
-3=probably false	3=probably true
-2=unlikely	2=likely
-1=somewhat unlikely	1=somewhat likely

0=no opinion

Problems to Estimate Posterior Probability with Evidence gathered

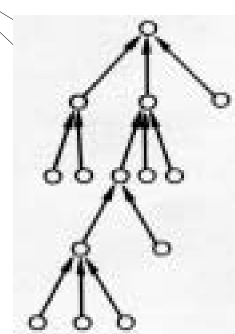
- The available evidence is generally incomplete and uncertain.
- The probabilistic relations link the hypotheses and relevant evidence are both unknown and complex.

Solution: Hierarchy Structuring

- The human expert will usually id 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.

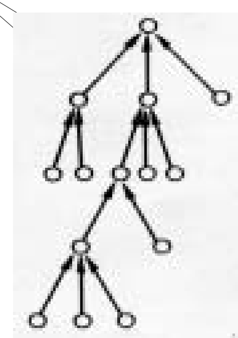
What Inference Networks Do

- 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.

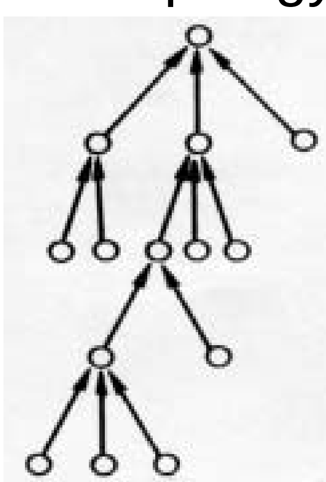


Categories of All assertions

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

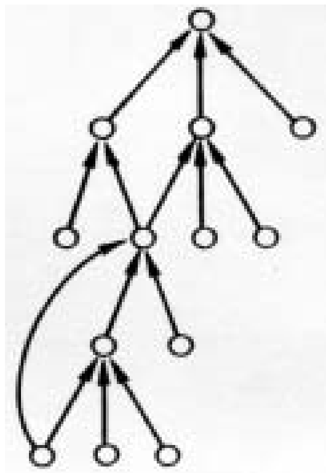


IN Topology(1) – Tree



If only one path from any evidence node to any top level hypothesis, the network has a tree structure.

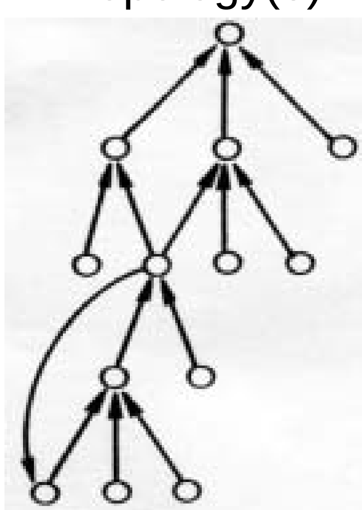
IN Topology(2) - Acyclic Graph



Multiple paths are not unusual. In this case the IN is a genuine graph.

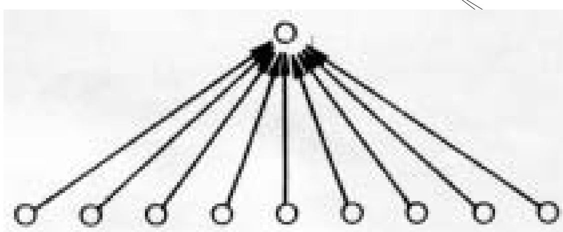
"Inference Networks are Acyclic Graph"

IN Topology(3) – Forbidden 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 … and A_k

$$P(A \mid E') = \min_i \{P(A_i \mid E')\}$$

- Disjunction
  $A=A_1$ or A_2 … or A_k

$$P(A \mid E') = \max_i \{P(A_i \mid E')\}$$

Combining Evidence: Weighted Combinations

- Prior Odds on A

$$O(A) = \frac{P(A)}{1 - P(A)}$$

- Likelihood Ratio (LR), "Sufficiency Measure"

$$\lambda_i = \frac{P(A_i \mid A)}{P(A_i \mid \overline{A})} \qquad ("LS")$$

- Bayes' rule states that:

$$\log O(A \mid A_1, A_2, \cdots, A_k) = \log O(A) + \sum_{i=1}^{k} \log \lambda_i$$

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:

$$\hat{\lambda}_i = \begin{cases} \lambda_i & if \quad P(A_i \mid E') = 1 \\ 1 & if \quad P(A_i \mid E') = P(A_i) \\ \overline{\lambda_i} & if \quad P(A_i \mid E') = 0 \end{cases}$$

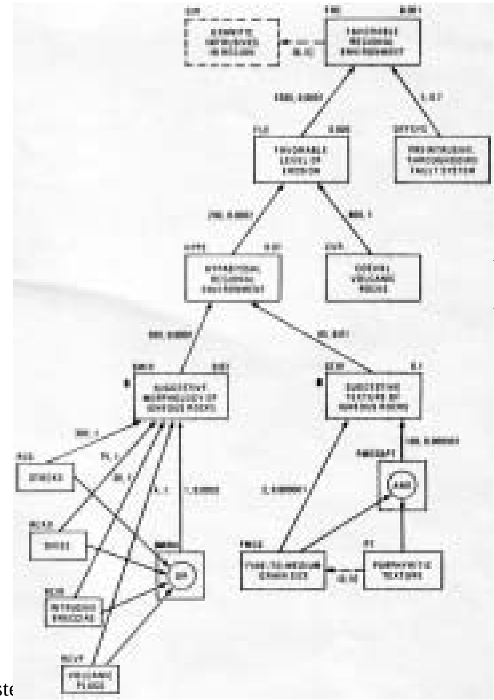
$\overline{\lambda}_i$ is the LR when A_i is known false, "Necessity Measure"

$$\overline{\lambda}_i = \frac{P(\overline{A}_i \mid A)}{P(\overline{A}_i \mid \overline{A})} \qquad ("LN")$$

Contexts and Subgoals

- Designate any proposition C as a *context*.
- Context arc (A→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*.

Inference Network for part of an ore deposit model



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.

Comparison of RWSSU model with designer for eight cases.

Test region	Target value	Prospector score	Difference
Black Hills	3.50	4.33	−0.83
Crooks Gap	4.70	4.26	0.44
Gas Hills	4.90	4.37	0.53
Shirley Basin	4.95	4.13	0.82
Ambrosia Lake	5.00	4.39	0.61
Southern Powder River	4.40	4.40	0.00
Fox Hills	1.50	2.17	−0.67
Oil Mountain	1.70	3.32	−1.62

Conclusions

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