UNIT 74 - KNOWLEDGE BASED TECHNIQUES

UNIT 74 - KNOWLEDGE BASED TECHNIQUES

Compiled with assistance from David Lanter, University of California, Santa Barbara

UNIT 74 - KNOWLEDGE BASED TECHNIQUES

Compiled with assistance from David Lanter, University of California, Santa Barbara

A. INTRODUCTION

- many geographical problems are ill-structured
- an ill-structured problem "lacks a solution algorithm and often even clear goal achievement criteria"

- goals are poorly defined
- data may be incomplete, or lack sufficient spatial resolution
- problem is complex - large volume of knowledge may be relevant to the problem
  - e.g. past experience with similar cases
  - e.g. precise knowledge in certain narrowly defined parts of the problem

- a DSS is one response to ill-structured problems
  - concentrates on delivering a wide range of functions to the user, rather than one solution
  - leaves the user with the role of expert

- knowledge based techniques are another
  - concentrate on making use of all available knowledge
  - goal is to emulate the reasoning of an expert
  - system takes the role of expert
  - the term "artificial intelligence" suggests the role of the machine in emulating the reasoning power of humans

Example

- where to put a label in a polygon? (the "label placement problem") - important in designing map output from GIS
  - goals are poorly defined - "maximize legibility", "maximize visual impact"
  - cannot turn goals into simple rules

- one rule might be "draw the label horizontally, centered at the centroid"
  - easy to turn the rule into an algorithm
  - rule is too simple - no good if the centroid lies outside the polygon - not clear how it affects legibility, visual impact
  - an expert system or knowledge based system should know when to use this rule, when not - may be many such rules

- there have been many attempts to reduce the label placement problem to a set of simple rules and build these into an "expert system"
  - ideally, the expert system could then perform the functions of a cartographer

Elements of knowledge based systems

- techniques for acquiring knowledge

- ways of representing knowledge internally
  - computers are good at representing numbers, words, even maps, but knowledge is potentially much more difficult

- search procedures for working with the internally stored knowledge

- inference mechanisms for deducing solutions to problems from stored knowledge

Expert system "shells"

- are software packages with functions which help the user construct special-purpose expert systems
  - provide a framework for organizing and representing knowledge
  - provide procedures for accessing knowledge in order to respond to queries or make decisions

- example applications of shells:
  - building a system to make medical diagnoses - emulating the medical expert
  - building a system to emulate the cartographer's knowledge of map projections, to pick the best projection for a particular problem

B. KNOWLEDGE ACQUISITION

- how is a knowledge base constructed?

- two approaches:
  - by asking experts to break their knowledge down into its individual facts, rules etc.
  - by deducing rules from the behavior of experts
  - both have been used in a GIS context

Example of knowledge base constructed by experts

- local government agency responsible for regulating land use in vast sparsely populated area - small staff

- must consider many hundreds of applications for land use permits annually, mostly from oil companies with large budgets and armies of lawyers

- decisions are subject to complex system of regulations, laws, past precedents, guidelines

- decisions must be defensible in court
  - desirable to know precise regulations, rules etc. which led to each decision
  - decisions must not be held to be arbitrary or capricious

- basic data - vegetation, soils, wildlife, geology etc. - in GIS

- knowledge base of all regulations, laws, precedents, guidelines

- decisions can be generated from knowledge base

Examples of knowledge inferred from interaction with experts

- Knowledge Based GIS (KBGIS) developed by Smith and others

- system can reduce query time by anticipating queries
  - e.g. certain overlay operations can be done in advance if the results will be needed frequently, redone when updates occur

  - e.g. certain topological relationships might be computed in advance and stored

- KBGIS analyzes queries received to "learn" about the pattern of queries and organize its database to optimize response
  - examines whether retrieving a stored fact takes longer than deducing it from other facts
  - if deducing it takes longer, the fact will be stored the first time it is deduced - subsequently it will be retrieved rather than deduced

- systems such as KBGIS learn about important spatial facts through the user's interaction with the system

C. KNOWLEDGE REPRESENTATIONS

- data structures in which knowledge can be stored

- more general than conventional databases

- four general methods for representing knowledge - trees, semantic networks, frames, production rules

Trees

- way of organizing objects that are related in a hierarchical fashion

- tree structures are common in geographical data
  - e.g. quadtrees and octrees
  - e.g. hierarchical nesting of census reporting zones

Semantic networks

- knowledge is organized as a set of nodes connected by labeled links
  - an algorithm can follow the links
  - e.g. topological data structures for road and river networks, boundaries of polygons (arcs)

- the GIS operations required to build an information product from input data layers can be visualized as a network of nodes and links
  - the links are GIS processes or functions, the nodes are datasets
  - this is a useful way of tracking the propagation of error through processes (links)

  - new datasets (nodes) inherit the inaccuracies of their predecessor datasets

Frames

- usually consist of the name of a phenomenon and the attributes that describe it
  - attributes are called "slots"

- increasing availability of frame based expert system shells

Production rules

- consist of two parts - situation part and action part
  - if situation exists, do the action
  - by convention left side is situation, right side is action

- most popular knowledge representation in geographical applications

- of the four areas of GIS - input, output, analysis and storage - output is most fully explored
  - production rules used in output for label placement, assignment of class intervals to choropleth maps, choice of projection
  - production rules for GIS analysis used in planning and resource management
  - production rules for GIS input center on scanning - rules for interpreting the image seen by the scanner, and vectorizing the image to create objects

D. SEARCH MECHANISMS

- need a procedure for accessing knowledge
  - "brute force" procedures test all knowledge contained in the database to obtain the best answer - only practical for small knowledge bases and simple problems
  - "heuristic" search procedures use rules designed to obtain the best answer or one close to it while minimizing search time

- each knowledge representation has associated search mechanisms
  - rules for searching trees dictate the branch to be taken at each fork
  - semantic networks are searched by examining the links at each node
  - frames - search for relevant frames, then relevant slots
  - for production rules, look for matching conditions on the left side of each rule

E. INFERENCE

- is the creation of new knowledge
  - the solution to any problem is new knowledge which can be stored in the system
  - a knowledge base can continue to grow as more knowledge is inferred from the existing base

  - e.g. a GIS can create new knowledge by computing topological relationships between objects from their geometrical relationships

- deductive inference:
  - creates new knowledge from existing facts through logical implication, e.g. using production rules
  - e.g. if A=B and B=C, then the system can deduce that A=C

- inductive inference:
  - produces new generalizations ("laws") which are consistent with existing facts
  - e.g. if the database contains the knowledge that area A is woodland and area B is woodland, and no information on any other area, the system might infer that all

areas are woodland

F. ISSUES

- knowledge based systems have been only moderately successful in areas where problems are relatively straightforward
  - e.g. medical diagnosis

- several factors may impede greater use:
  - high cost of developing system - building the knowledge base
  - uniqueness of every application
  - dynamic nature of knowledge - knowledge base is not static
  - inadequacy of alternatives for knowledge representation - few examples fit precisely within any one form, e.g. production rules
  - unwillingness to trust the decisions of a machine (no "bedside manner")
  - response time deteriorates rapidly as knowledge base grows
  - most knowledge is "fuzzy" or uncertain - system must return many possible answers to a problem - few problems have a precise, single answer - technical difficulties of representing and processing fuzzy knowledge
  - poor design of user interface - not "user friendly"
  - user often wants the reasoning behind a decision, not just the decision itself

- some of the most successful applications have been for instruction
  - e.g. use of medical expert system to develop diagnostic skills - encouraging students to structure knowledge and process it systematically in response to a problem

  - as precise, analytical models of knowledge and the ways in which it is used, expert systems can enhance our understanding of human decision-making processes - e.g. how does a cartographer position labels on a map?

REFERENCES

Texts on artificial intelligence and expert systems:

Luger, G.F. and W.A. Stubblefield, 1989. Artificial Intelligence and the Design of Expert Systems, Benjamin/Cummings Publishing Co, Redwood City, CA.

Tanimoto, S.L., 1987. The Elements of Artificial Intelligence, Computer Science Press, Rockville, MD.

Winston, P.H., 1980. Artificial Intelligence, Addison-Wesley, Reading, MA.

KBGIS:

Smith, T.R. and M. Pazner, 1984. "Knowledge-based control of search and learning in a large-scale GIS," Proceedings, International Symposium on Spatial Data Handling, Zurich, 2:498-519.

Smith, T.R. et al., 1987. "KBGIS-II: a knowledge-based geographical information system," International Journal of Geographical Information Systems 1:149-72.

Other:

Freeman, H. and J. Ahn, 1984. "AutoNAP an expert system to automate map name placement," Proceedings, International Symposium on Spatial Data Handling, Zurich, pp 556-571. Design of an expert system for polygon label placement.

Imhof, E., 1975. "Positioning names on maps," The American Cartographer 2. An analysis of rules for label positioning.

Kubo, S., 1986. "The basic scheme of TRINITY: a GIS with intelligence," Proceedings, Second International Symposium on Spatial Data Handling, Seattle. International Geographical Union, Commission on Geographical Data Sensing and Processing, Williamsville, NY, 363-74.

Walker, P.A. and D.M. Moore, 1988. "SIMPLE: an inductive modelling and mapping system for spatially oriented data," International Journal of Geographical Information Systems 2:347-63.

EXAM AND DISCUSSION QUESTIONS

1. Compare the use of knowledge bases and inference in Smith's KBGIS, Kubo's TRINITY and Walker and Moore's SIMPLE. What general principles of knowledge based systems do they each exploit? Which application do you consider the most successful?

2. Artificial intelligence has often been called the study of a set of unsolved problems. However, once an algorithm has been devised to solve a given problem, it becomes simply a solved problem, no longer meriting the mystique associated with the term "artificial intelligence". Do you agree?

3. What areas of GIS - applications, input techniques, processes etc. - do you consider most suitable for development of expert systems?

4. Discuss the differences between spatial decision support systems and knowledge based systems as alternative approaches to solving poorly structured problems.

Last Updated: August 30, 1997.