1187

TRANSPORTATION RESEARCH RECORD

Expert Systems in Transportation

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1988

Transportation Research Record 1187
Price: $9.00
Editor: Kieran P. O'Leary

modes
1 highway transportation
2 public transit

subject areas
12 planning
21 facilities design

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Expert systems in transportation.
(Transportation research record, ISSN 0361-1981 ; 1187)
1. Highway engineering–Data processing. 2. Expert systems (Computer science) I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 1187 [TE153] 380.5 s 89-12930
ISBN 0-309-04753-6 [388'.0285'633]

Sponsorship of Transportation Research Record 1187

GROUP 5 INTERGROUP RESOURCES AND ISSUES
Chairman: William M. Spreitzer

Task Force on Expert Systems
Chairman: Roswell A. Harris, University of Louisville
Domenick J. Billera, William Bowlby, Anselmo Osvaldo Braun, Edmond Chin-Ping Chang, Louis F. Cohn, Michael J. Demetsky, Jon D. Fricker, Geoffrey D. Gosling, Jerry J. Hajek, Chris T. Hendrickson, Mike Kaul, Olin W. Mintzer, Satish Mohan, Peter S. Parsonson, Michel Ray, Stephen G. Ritchie, Stephen Edwin Rowe, Alberto J. Santiago, Jerry B. Schneider, John R. Stone, George F. Swetnam, Hossein B. Takallou, James A. Wentworth, Marcus Ramsay Wigan

James A. Scott, Transportation Research Board staff

The organizational units, officers, and members are as of December 31, 1987.

Transportation Research Record 1187

Contents

Foreword

Faghri and Demetsky examine the significance of the means of representing the knowledge base in the development of expert systems, with special reference to transportation engineering. The development, highlights, and shortcomings of each representation technique are discussed as well as the results of expert systems and how they relate to different representation techniques.

In the paper by Adams et al., the authors describe the architecture and several modules of an expert system aid for retaining wall rehabilitation design. The system could be used to evaluate retaining wall failure, survey the condition of existing walls, or design rehabilitation or new construction strategies.

Ritchie et al. present the results of a research project in their paper. The major objective of the project was to prepare a plan for the development and implementation of a knowledge-based expert systems project throughout the California Department of Transportation (Caltrans). A total of 45 candidate projects were identified and ranked by priority. Caltrans has now begun implementation of the plan developed in this research.

Janarthanan and Schneider discuss an interactive knowledge-based expert system (TNOP_ADVISOR) that was developed to assist in the development of high-performance transit network designs. The system described provides advice about how to modify designs to obtain improved performance. A network software package provides the capability for modifying and predicting the performance of the designs.

Braun and Machado de Sá discuss the research approach to development of contingency planning being performed by the Rio de Janeiro State Department of Transportation. Completion of operational plans for recovering transportation services in cases of disruption caused by emergency or contingency will be followed by the development of corresponding expert systems to assist in decision making.

Knowledge Representation and Software Selection for Expert Systems Design

ARDESHIR FAGHRI AND MICHAEL J. DEMETSKY

A variety of techniques and methods for representing knowledge in the knowledge base of expert systems have been used. The authors examine the significance of the means of representing the knowledge base in the development of expert systems, with special reference to transportation engineering. The development, highlights, and shortcomings of each representation technique are discussed, and appropriate transportation engineering examples are given. Also presented are the results of an investigation of expert system tools and how they relate to different representation techniques.

The heart of an expert system is its knowledge, which is structured to support decision making. When scientists in artificial intelligence (AI) use the term "knowledge," they mean the information a computer needs before it can function intelligently (*1*); this information takes the form of facts and rules. Facts are truths in some relevant world—things we want to represent. Representations of facts are the things we will actively be able to manipulate. For example,

Fact: Responses to a brake light from a leading vehicle require 0.4 sec to more than 1.0 sec for some drivers (*2*).

Example: All physical motor capabilities deteriorate with age.

Rules are formal representations of recommendations, directives, and strategies; they may be expressed as conditional (if-then) statements. For example,

Rule: If forced flow and low speeds exist on a segment of highway, a level of service F is achieved.

Example: If the degree of congestion or vehicle delay, or both, caused by daytime lane closures is severe, nighttime construction and maintenance should be considered.

Facts and rules in an expert system are not always true or false; sometimes there is a degree of uncertainty about the truth of a fact or the validity of a rule. When this doubt is made explicit, it is called a certainty factor (*1*).

Fact: Rail–highway crossings near major employment centers experience more accidents with certainty 0.7.

Rule: If the average speed increases by 10 mph with certainty 1.0, the number of accidents will increase by 10 percent with certainty 0.6.

The organization of knowledge in an expert system separates the knowledge about the problem domain from the system's other knowledge, such as general knowledge about how to solve problems or knowledge about how to interact with the user. The collection of domain knowledge is called the knowledge base; the general problem-solving knowledge is called the inference engine (*1*).

Different techniques and methods for representing knowledge in the knowledge base of expert systems have been used. The authors examine the significance of the means of representing the knowledge base in the development of expert systems, with special reference to transportation engineering. The development, highlights, and shortcomings of each representation technique are discussed, and appropriate transportation engineering examples are given. Also presented are the results of a thorough investigation of expert system tools and how they relate to different representation techniques.

KNOWLEDGE REPRESENTATION TECHNIQUES

In expert systems, complex problem solving requires both a large amount of knowledge and a mechanism for manipulating that knowledge to create solutions to new problems. A number of methods for representing knowledge (facts) have been used in expert systems. In this paper, the common knowledge representation techniques are discussed: predicate logic, other logics, structured representation, rules, and object-attribute-value triplets.

Predicate Logic

Logic is critically concerned with the validity of arguments, that is, methods of determining whether given conclusions can be validly drawn from given facts. Logic is relevant to programming because a program is really a set of quasi-logical statements that are processed in some way to generate a conclusion (*3*). In logic a "true argument" has a

Virginia Transportation Research Council, Box 3817, University Station, Charlottesville, Va. 22903.

precise, clearly defined meaning: an argument is considered true if and only if all of its assumptions are true; then its conclusions are true also.

To decide on the acceptability of a particular argument, it is necessary to make some test. In logic the method of doing this is to compare the text of interest with abstracted patterns of argument to seek a match. Such patterns are termed "forms" and are made up of abstracted sequences of facts and rules that have been proved valid in a mathematical (or "formal") way (*3*).

The capability of logic to generate (or infer) new information from old is of particular interest given the tendency to view programming as the controlled generation of inferences. Moreover, with many years of development behind it, logic also provides a well-defined and well-understood formalism for representing facts and the rules for manipulating them.

Before discussion of the concepts and applications of predicate logic, using propositional logic as a way of representing the sort of world knowledge an expert system might need is explored. Propositional logic is appealing because it is simple to deal with and there is a decision procedure for it. Real-world facts can easily be represented as logical expressions (or logical propositions) written as well-formed formulas (wff's) in propositional logic, such as the following (*4*):

It is raining.
RAINING
It is sunny.
SUNNY

These propositions could be used, for example, to deduce that it is not sunny if it is raining:

If it is raining then it is not sunny.
RAINING -SUNNY

But it is easy to observe the limitations of propositional logic. The obvious fact stated in "an automobile is a vehicle" could be written "Autovehicle." But "a truck is a vehicle" would have to be written "Truckvehicle."

This would be a totally different assertion, and no conclusions could be drawn about similarities between "auto" and "truck." It would be much better to represent these facts as "Vehicle (auto)" and "Vehicle (truck)," because the structure of the representation would reflect the structure of the knowledge itself. It is even more difficult if we try to represent "all vehicles are unsafe," because quantification would be needed unless separate statements were written about the safety of every known vehicle.

So it appears that predicate logic must be the way to represent knowledge, because it permits representations of things that cannot reasonably be represented with propositional logic. In predicate logic, real-world facts can be represented as statements written as wff's. But a major motivation for choosing to use logic was that if logical statements were used as a way of representing knowledge, then there would be a good way to reason with that knowledge. Determining the validity of a proposition in propositional logic is straightforward, though computationally it may be difficult.

Predicate logic provides a way of deducing new statements from old ones. Unfortunately, however, unlike propositional logic, it does not have an associated decision procedure (*4*). There are procedures that will lead to a proof of a proposed theorem (if indeed it is a theorem), but they will not necessarily halt if the proposed statement is not a theorem. One such simple procedure is to use the rules of inference to generate theorems from axioms in some orderly fashion, testing each to see if it is the one for which a proof is sought. This method, however, is not very efficient, and investigation continues to find better ones. So, despite the theoretical undecidability of predicate logic, it can still serve as a useful way of representing and manipulating some of the kinds of knowledge that an expert system might need.

Knowledge representation by using predicate logic is demonstrated by the examples shown below. Consider the following statements:

1. Lee Highway was congested.
2. Lee Highway is an Interstate.
3. All Interstates are highways.
4. Washington Metro is a heavy rail train.
5. All people either like downtown New York or hate it.
6. People only try to ignore freeways that are congested.

The facts described by these sentences can be represented as a set of wff's in predicate logic. But some notation must be defined first:

$\wedge$ = AND
$\vee$ = OR
$-$ = negation
$\forall$ = for every
$\exists$ = there exists

By using this notation, the predicate logic version of the six statements may be presented as follows:

1. Lee Highway was congested.
   Congested (Lee Highway)

This representation captures the critical fact that Lee Highway is congested. It fails to capture some of the information in the English sentence, namely, the notion of past tense. Whether this omission is acceptable depends on how the knowledge is to be used.

2. Lee Highway is an Interstate.
   Interstate (Lee Highway)
3. All Interstates are highways.
   $\forall$ xInterstate(x) $\rightarrow$ Highway(x)
4. Washington Metro is a heavy rail train.
   Heavy rail train (Washington Metro)

Here we ignore the fact that proper names are often not references to unique items, because many things share the same name. Sometimes deciding which of several things is being referred to in a particular statement may require a fair amount of knowledge and reasoning.

5. All people either like downtown New York or hate it.

$\forall$ xPerson(x) $\rightarrow$ like(x, New York) $\vee$ hate(x, New York)

In English, "or" sometimes means the logical inclusive "or" and sometimes means the logical exclusive "or." Here the inclusive "or" is used. Some may argue that this English sentence is really stating an exclusive "or." Its expression would be:

$\forall$ xPerson(x) $\rightarrow$ [like(x, New York) $\vee$ hate(x, New York)]

$\wedge \sim$ [like(x, New York) $\wedge$ hate(x, New York)]

6. People only try to ignore freeways that are congested.

$\forall$ x $\forall$ y person(x) $\wedge$ freeway(y) $\wedge$ tryignore(x, y)

This sentence, too, is ambiguous. Does it mean that the only freeways that people try to ignore are those that are congested (the interpretation used here)? or Does it mean that the only thing people try to do is to ignore congested freeways?

From these statements, three important issues must be addressed when converting English sentences to logical statements and then using those statements to deduce new ones:

1. Many English sentences are ambiguous. Choosing the correct interpretation may be difficult.
2. There is often a choice of ways of representing the knowledge. Simple representations are desirable, but they may preclude certain kinds of reasoning. The useful representation for a particular set of sentences depends on the use to which the knowledge contained in the sentences will be put.
3. Even in very simple situations, a set of sentences is unlikely to contain all the information necessary to reason about the topic at hand.

Although predicate logic is a useful way of representing knowledge for many expert system domains, some kinds of information may not be easily represented by this method. Discussed in the next section are other useful methods of knowledge representation.

Other Logics

The techniques of predicate logic are useful for solving problems in many different domains. But unfortunately in many other interesting domains, predicate logic does not provide a good way of representing and manipulating important information. Such domains are mostly uncertain and fuzzy. The methods discussed in this section for these problems with computer programs are monotonic logic and statistical and probabilistic reasoning.

Monotonic Logic

Monotonic logic allows statements to be deleted from as well as added to the data base. Among other things, it allows belief in one statement to depend on lack of belief in some other one. Rarely does a system contain all the information that would be useful. But often when such information is lacking, some sensible guesses can be made as long as there is no contradictory evidence. The construction of these guesses is known as "default reasoning."

For example, suppose the chief traffic engineer of a large metropolitan city decides to add one lane to a road in the city's street network that has been determined to have severe congestion during peak hours (after all economic and social issues have been resolved). Would adding a lane relieve the congestion? The engineer can approach the problem fairly well if he uses a general rule: Because the peak-hour volume exceeds the capacity of the facility in question, adding an extra lane would relieve the congestion unless there is evidence to the contrary.

This sort of default reasoning is nonmonotonic (i.e., the addition of one piece of information may force the deletion of another), because statements so derived depend on lack of belief in certain other statements. This means that if one of those previously lacking statements is added to the system, the statement generated by default reasoning will have to be deleted. Thus, in our example, if the engineer fails to realize that the fundamental differences between the normative system optimization (SO) flow pattern and the descriptive user equilibrium (UE) flow pattern on the network may lead to Braess's paradox (*5*) (the addition of the lane may actually increase the travel time of the vehicles), he should delete his previous belief that this particular strategy will work. Of course, he must also delete any other beliefs that are based on the belief that has just been discarded. This kind of default reasoning is referred to as the "most probable choice" (*4*).

In general, nonmonotonic reasoning systems may be necessary because of (a) incomplete information, which requires default reasoning; (b) changing knowledge that must be described by a changing data base; or (c) a complete solution to problems, which may require assumptions about partial solutions.

Statistical and Probabilistic Reasoning

In representing the knowledge in expert systems, it is assumed that either a fact is known to be true, or it is known to not be true, or nothing at all is known about it. Still to be considered is the possibility of facts that may be "probably true." There are three kinds of situation in which probabilistic reasoning may be employed:

1. The relevant world is really random, for example, the motion of electrons in an atom or the distribution of speeds on a certain highway.
2. The relevant world is not random given enough data, but a system will not always have access to that many

data, for example, the likelihood of success of a traffic control strategy to combat congestion.

3. The world appears to be random because it has not been described at the proper level.

Probabilistic reasoning in the first two cases is utterly appropriate. The mathematical theory of probability provides a way of describing and manipulating uncertain knowledge. Sometimes very simple techniques of probability can be used effectively in expert systems.

One of the most useful results of probability theory is Bayes's theorem, which provides a way of computing the probability of a particular event given some set of observations. The theorem states:

$$P(H_i \mid E) = \frac{P(E \mid H_i) * P(H_i)}{\sum_{n=1}^{k} P(E \mid H_n) * P(H_n)}$$

where

$P(H_i \mid E)$ = probability that hypothesis i is true given evidence E,
$P(E \mid H_i)$ = probability that evidence E will be observed given that hypothesis i is true,
$P(H_i)$ = a priori probability that hypothesis i is true in the absence of any specific evidence, and
k = number of possible hypotheses.

Bayes's theorem can be modified to handle a variety of more complicated situations. For example, a single body of evidence E might not be collected all at once. Rather, a series of smaller observations might be made over time. Other results in probability theory can also be applied to these kinds of problems.

Structured Representation

A good system of representing complex structured knowledge in a particular domain for use in expert systems should have the following four properties (*4*):

1. Representation adequacy—the ability to represent all of the kinds of knowledge that are needed in that domain.
2. Inferential adequacy—the ability to manipulate the representational structures in such a way as to derive new structures corresponding to new knowledge inferred from old.
3. Inferential efficiency—the ability to incorporate into the knowledge structure additional information that can be used to point the inference mechanisms in the most promising directions.
4. Acquisitional efficiency—the ability to acquire new information easily. The simplest case involves direct insertion of new knowledge into the data base. Ideally, the program itself would be able to control knowledge acquisition.

The representation techniques discussed previously are useful for representing simple facts, but they cannot always have the desired properties of a representation technique. Several techniques for acquiring these properties have been developed. These techniques are referred to as "declarative methods" (*4*). In declarative knowledge representation, most of the facts are presented as a static collection of knowledge accompanied by a small set of general procedures for manipulating them. In this section, three declarative mechanisms for representing knowledge are presented: semantic nets, frames, and scripts.

Semantic Nets

The term "semantic net" is used to describe a knowledge representation method that is based on a network structure. Semantic nets were originally developed for use as psychological models of human memory but are now a standard method of representation for artificial intelligence and expert systems. A semantic net consists of points (nodes) connected by links (arcs) describing the relations between the nodes. The nodes in a semantic net represent objects, concepts, or events. Arcs can be defined in different ways, depending on the kind of knowledge being represented.

Isa arcs are most often used to establish a property inheritance hierarchy; that is, instances of one class have all properties of more general classes of which they are members. Has-part arcs identify nodes that are properties of other nodes. Figure 1 shows both isa and has-part arcs in a simple net for the concept of a public transit mode.

The isa relation, like the has-part relation, establishes an inheritance hierarchy for properties in the net (*1*), so that items lower in the net inherit properties from items higher in the net. This saves space, because information about similar nodes does not have to be repeated at each node and can be stored in one central location. For example, in the public transit-mode semantic net the common parts of each node, such as passenger seats and engine, are stored once at the node level instead of repeatedly at lower levels like a bus or a particular bus system. The net can be searched, by using knowledge about the meaning of the relations in the arcs, to establish facts like "Washington Metro has passenger seats." Semantic nets are a useful way to represent knowledge and to simplify problem solving in domains that use well-established taxonomies (*1*).

Frames

Frames provide another method of representing facts and relationships. A frame is a description of an object that contains slots for all the information associated with the object. Values may be stored in slots. Each slot can have any number of procedures attached to it. Three useful kinds of procedure often attached to slots are

1. If-added—executes when new information is placed in the slot,

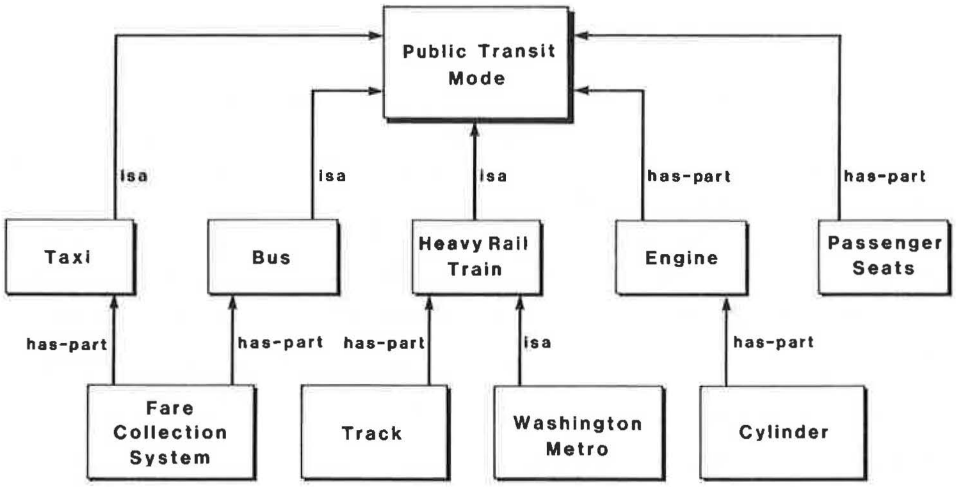

FIGURE 1 Semantic network, showing isa and has-part arcs.

2. If-removed—executes when information is deleted from the slot, and
3. If-needed—executes when information is needed from the slot but the slot is empty.

These attached procedures can monitor the assignment of information to the node, thereby ensuring that appropriate action is taken when values change.

A frame is organized much like a semantic net. It is a network of nodes and relations organized in a hierarchy in which the higher nodes represent general concepts and the lower nodes represent properties of those concepts. Frame systems are useful for problem domains in which expectations about the form and content of the data play an important role in problem solving, such as interpreting visual scenes or understanding speeches. Figure 2 shows an example of a frame network for an expert system.

Scripts

A script is a structure in which a stereotyped sequence of events in a particular context is described. A script consists of a set of slots. Associated with each set of slots may be some information about what kind of values it may contain, as well as a default value to be used if no other information is available. So far this definition seems similar to that for frames, but scripts have other important components, a few of which are

1. Entry conditions—conditions that, in general, must be satisfied before the events described in the script can occur;
2. Results—conditions that, in general, will be true after the events described in the script have occurred;
3. Props—slots that represent objects that are involved in the events described in the script (the presence of these objects can be inferred even if they are not mentioned explicitly);
4. Roles—slots that represent people who are involved in the events described in the script (the presence of these people, too, can be inferred even if they are not mentioned explicitly—if specific individuals are mentioned, they can be inserted into the appropriate slots); and
5. Track—the specific variation on a more general pattern that is represented by the particular script (different tracks of the same script will snare many but not all components).

Although scripts are less general than are frames, and so are not suitable for representing all kinds of knowledge, they can be very effective for representing the specific kinds of knowledge for which they are designed.

Rules

In expert systems the term "rule" refers to the most popular type of knowledge representation technique, the rule-based representation. Rules provide a formal way of representing recommendations, directives, and strategies; they are often appropriate when the domain knowledge results from empirical associations developed through years of problem-solving experience. Rules are generally expressed as conditional (if-then) statements. Rules might exist in an expert system for determining whether to rehabilitate or replace highway bridges:

1. If a bridge has a sufficiency rating between 50 and 80, then it should be rehabilitated.
2. If a bridge is scheduled to be replaced within 6 yr, then only routine maintenance will be necessary until it is replaced.
3. If any one component of a bridge (namely, the substructure, superstructure, or deck) has a condition rating greater than 5 and the bridge is less than 20 years old, than only routine maintenance will be required on that component.

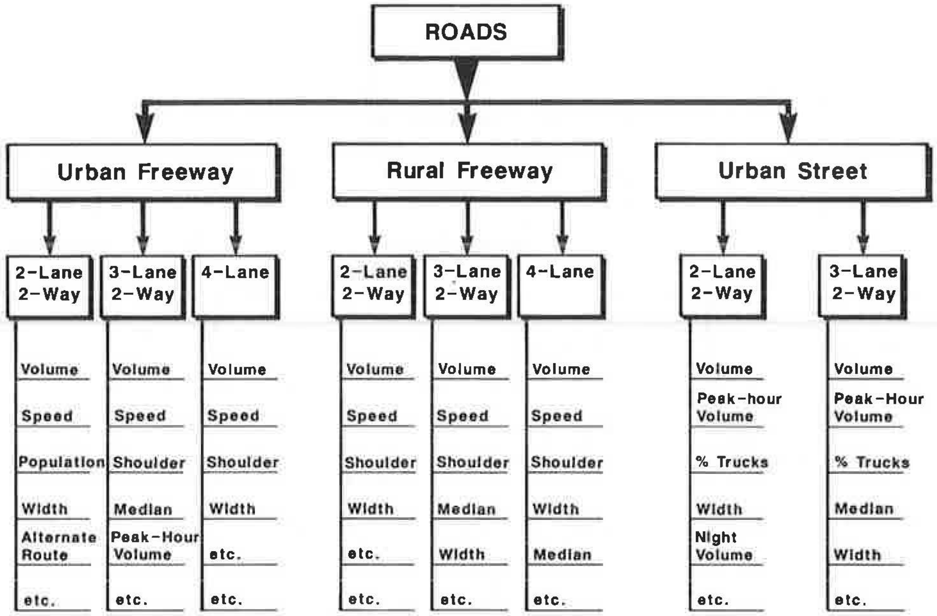

FIGURE 2 Frame network.

Each of the two parts of the antecedent in Rule 3 is called an "expression," or "if clause." The consequent usually contains a single expression, or "then clause"; it could contain more than one. The clauses in the antecedent can be connected by the logical operator "and" or "or."

In a rule-based expert system, knowledge is represented as sets of rules that are checked against a collection of facts or knowledge about the current situation. When the antecedent of a rule is satisfied by the facts, the action specified by the consequent is performed. When this happens, the rule is said to "fire" or "execute" (*1*). A rule interpreter compares the antecedents with the facts and executes the rule whose consequent matches the facts, as follows:

Facts: A flammable liquid was spilled.
The pH of the spill materials is less than 6.
The spill smells like vinegar.
Rule: If the pH of the spill is less than 6, the spill material is acid.

The new fact is added to the knowledge base: The spill material is an acid.

The action of the rule may modify the set of facts in the knowledge base by adding a new fact. The new facts added to the knowledge base can themselves be matched to the antecedent of the rule. The matching of rule antecedents to the facts can produce what are called "inference chains" (*1*). The inference chain for this example is shown in Figure 3. This inference chain shows how the system used the rules to infer the identity of the spill material. An expert system's inference chains can be displayed to the user to help explain how the system reached its conclusions.

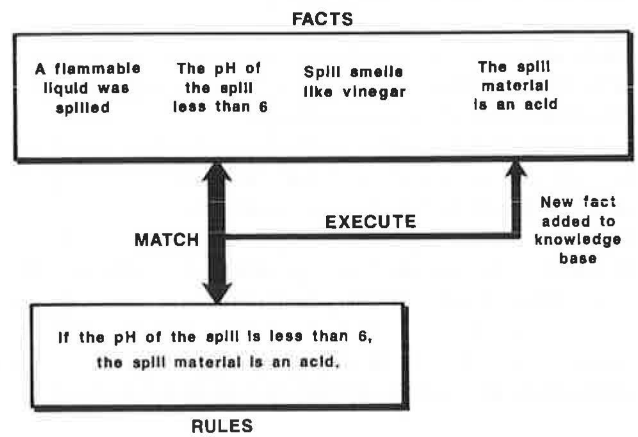

FIGURE 3 Inference chain.

Object-Attribute-Value Triplets

Another way to represent factual information is by object-attribute-value (OAV) triplets. In this scheme, an object may be a physical entity such as a door, a car, or a pavement, or it may be a conceptual entity, such as a logic gate, a bank loan, or a sale. An attribute is a general characteristic or property of an object; for example, interest rate is an attribute of a bank loan. The final member of the triplet is the value, which describes the specific nature of an attribute in a particular situation. For example, the number of lanes on a certain highway might be 6, or the interest rate for a bank might be 12 percent. Figure 4 shows an example of OAV representation.

Representing knowledge with OAV triplets is a specialized form of semantic network. Exotic links are banished in favor of two simple relationships. The object-attribute

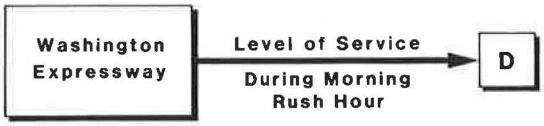

FIGURE 4 OAV representation.

link is a has-a link, and the attribute-value link is an isa link. For example, a bank loan has a rate of interest, and 12 percent is a rate of interest. Nodes are classified as objects, attributes, or values.

EXPERT SYSTEM TOOLS

Expert system tools are the programming languages and support packages used to build the expert system. The three major categories of tools available for expert system building are programming languages, knowledge engineering languages, and system building aids. The most common tools currently used by transportation engineers to develop expert systems are knowledge engineering languages (SHELLs) because they are relatively easy to use. However, depending on the nature of the problem and the kind of representation a knowledge engineer chooses, a different kind of expert system building tool may be employed. The three categories of expert system tools, some of the current commercial systems available in each category, and the kind of representation technique for which each system was designed are described next.

Programming Languages

The programming languages used in expert systems applications are generally either problem-oriented languages, such as FORTRAN and Pascal, or symbol-manipulation languages, such as LISP and Prolog. Currently, the most popular symbol-manipulation language for expert system applications is LISP. A feature of LISP that distinguishes it from most other languages is its mechanism for manipulating symbols. LISP can manipulate symbols readily because of its list structure characteristics. List structures are collections of items enclosed in parentheses, in which each item can be either a symbol or another list. Complex concepts can be represented in and built into an expert system using the list structures.

Problem-oriented languages are generally designed for solving particular classes of problems. FORTRAN, for example, performs algebraic calculations for scientific, mathematical, and statistical problems. Problem-oriented languages have been used in expert system development but are not very popular for extensive applications. Some of the commercial programming languages for developing expert systems are INTERLISTP-d and SMALLTALK-80 (Xerox Corporation); LISP (LISP Machine, Inc.); Prolog (Quintus Computer Systems, Inc.); and GCLISP (Gold Hill Computers, Inc.).

Knowledge Engineering Languages (SHELLs)

Knowledge engineering languages are a subclass of programming languages designed specifically for expert systems development. They fall into two major categories: skeletal systems and general-purpose systems. Removing from the expert system domain-specific knowledge leaves the skeletal system, the inference engine, and the support facilities. Support facilities are the environment associated with a building tool in the expert system that helps the user interact with it. Because skeletal systems apply only to a limited class of problems, they lack generality and flexibility as a building tool method. The structure and built-in facilities of a skeletal system, however, make expert systems development easy and fast. The key decision that must be made initially by the system developer is to select an appropriate SHELL that matches the problem.

In contrast, general-purpose knowledge engineering languages can handle a wide range of problem areas and types. They provide more control over accessing information in the knowledge base than does a skeletal system.

The general-purpose languages, however, may be more difficult to use. Table 1 shows some commercially available SHELLs along with the type of representation technique for which they were designed.

System-Building Aids

System-building aids consist of commercially available software programs that can be classified as either design aids or knowledge acquisition aids. Design aids help the expert system developer design and build an expert system by establishing a framework for the representation of knowledge and its supporting facilities. Knowledge acquisition aids help the expert system builder transfer the knowledge rules and heuristics from the human expert to the knowledge base of an expert system. Available expert system building aids include EXPERT-EASE (Expert Systems International), RULE-MASTER (Radian Corporation), and TIMM (General Research Corporation).

TABLE 1 SELECTED EXPERT SYSTEMS SHELLs

Tool	Representation	Developer
ART	Rule and frame-based	Inference Corporation
DUCK	Logic and rule-based	Smart Systems Technology
EXSYS	Rule-based	Exsys, Inc.
KEE	Rule and frame-based	Intellicorp
M.I.	Rule-based	Teknowledge
OPS 5	Rule-based	Digital Equipment Corporation
S.1	Rule and frame-based	Teknowledge
SRL+	Frame-based	Carnegie Graphics Inc.

CONCLUSION

Presented in this paper are the options available in expert systems technology for building systems to aid in solutions for transportation engineering problems. First, the problem must be defined in terms of the most appropriate way for source knowledge about the problem to be represented. Eight common techniques for representing knowledge are: predicate logic, monotonic logic, statistical and probabilistic reasoning, semantic nets, frames, scripts, rules, and object-attribute-value. Rules are the most commonly used because of the availability of rule-based SHELLs. Second, the problem must be matched to a practical system building tool. When these major decisions are made, the knowledge acquisition and system development process can proceed. The state of the practice of building expert systems will mature when categories of tools can be related to classes of problems that generalize categories of expert systems applications in transportation engineering.

ACKNOWLEDGMENT

This paper was produced with the cooperation of the FHWA, U.S. Department of Transportation.

REFERENCES

1. D. A. Waterman. *A Guide to Expert Systems.* Addison-Wesley Publishing Company, Reading, Pa., 1986.
2. E. C. Carter and W. S. Hamburger. *Introduction to Transportation Engineering.* Reston Publishing Company, Inc., Reston, Va., 1978.
3. J. L. Alty and M. J. Coombs. *Expert Systems.* NCC Publications, Manchester, England, 1984.
4. E. Rich. *Artificial Intelligence.* McGraw-Hill Book Company, New York, 1983.
5. Y. Sheffi. *Urban Transportation Networks.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1985.

Expert System Architecture for Retaining Wall Design

T. M. Adams, C. Hendrickson, and P. Christiano

The design of emergency repairs and the growing inventory of failing or newly repaired retaining walls present a considerable practical problem. For repair work, time and budget constraints are pressing. Available expertise is limited and often dispersed among numerous individuals, so extensive consultation is often required. If an intelligent aid were available, engineers and managers would be able to make better and faster decisions. Described in this paper are the architecture and several modules of an expert system aid for retaining wall rehabilitation design. Such a system could be used by an engineer who is called upon to evaluate a retaining wall failure, conduct a survey of the condition of existing walls, or design a rehabilitation or new construction strategy for a wall. Issues addressed include appropriate problem paradigm models, alternative solution strategies, and reasoning with uncertainty.

Preliminary design and conceptual cost estimation are fundamental engineering activities in the provision of newly constructed facilities or in the repair and rehabilitation of existing infrastructure. Initial decisions at the conceptual or planning stage form the basis for detailed engineering design and construction agreements. Poor decisions can cause delays, excessive costs, or, if mistakes are not recognized, unsafe and inadequate facilities. Given the importance of this activity to the nation's infrastructure, improvement of the practice of preliminary design and conceptual cost estimation is of great value.

Retaining walls provide an excellent example of the problems and processes associated with preliminary engineering design and cost estimation. These walls are ubiquitous in the United States and perform as necessary components of a variety of infrastructure facilities including roads, bridges, buildings, drainage systems, and so on. In essence, retaining walls act as "infrastructure components" for infrastructure systems. Repairing, replacing, and rehabilitating the large number of existing walls are common engineering problems involving consideration of numerous options, such as simple concrete crack repairs, replacement of reinforcing steel, or insertion of tiebacks. The option of fully replacing damaged walls also raises a variety of alternatives for overall design of new walls and problems of stabilizing existing soil conditions during construction. Assessing the technical feasibility, cost, and overall desirability of different options requires considerable judgment and expertise by the engineer or manager.

Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh, PA 15213.

Most designers of retaining walls are skillful at determining a nearly optimal configuration of a particular design once the decision has been made to construct a wall of a particular type. If a reinforced concrete cantilever wall has been selected, for example, standard design procedures lead to the choice of concrete thickness and amount of reinforcing of the stem, as well as the dimensions and reinforcing of the toe, heel, and key. An individual engineer may find it much more difficult, however, to determine whether a cantilever wall is the best choice among those available. In the last few decades, a smaller fraction of traditional cast-in-place-concrete gravity and cantilever walls has been built because of a rising popularity of earthwork reinforcement and prefabricated modular systems (*1*). In many situations, specialty proprietary systems can be constructed at 30 to 50 percent less cost and construction time than for conventional cast-in-place walls (R. M. Leary and G. L. Klinedinst, unpublished data). The issue becomes even more important for walls having heights greater than 10 ft as the relevant number of design alternatives increases. As a result, considering numerous alternative wall systems is required for acceptable engineering practice.

The task of determining feasible designs for prescribed conditions is not easy, because many different retaining systems are available. However, project constraints and site conditions often dictate the type of retaining structure that should be considered. Some wall systems may be eliminated because of unavailability of materials, necessary service life, or environmental and aesthetic requirements. Special loading requirements, anticipated settlements, and adaptability to field changes must also be considered before a wall system is chosen. To make an intelligent decision requires familiarity with the available alternatives. Furthermore, a comparison of complete wall designs is often necessary before selection of one that will perform satisfactorily and can be constructed at the lowest overall cost.

To help state highway departments choose and contract alternative retaining wall systems, some guidelines have been established by the FHWA. However, the recent focus of state highway projects has shifted from new construction to reconstruction and widening of existing facilities (R. M.

Leary and G. L. Klinedinst, unpublished data). Thus the problems associated with economically and effectively dealing with old and failing retaining walls are becoming relevant. As another example, the city of Pittsburgh has hundreds of walls that were built at the turn of the century, many of which need immediate rehabilitation or replacement (*2*).

Knowledge-based expert systems (KBES) provide a practical means to incorporate experts' heuristics, rules of thumb, judgments, empirical associations, and historical information into a computer program that can assist and guide practitioners. In contrast to algorithmic processes, KBES need not have all contingencies preprogrammed and anticipated. Although these systems have already been applied in a variety of applications (*3, 4*), the use of expert systems technology in infrastructure rehabilitation has not been extensively investigated.

In this paper, the expert systems architecture for a retaining wall design assistant is described. A prototype system called RETAIN has been implemented in C language and the OPS5 expert system environment. The RETAIN project focuses on categorizing and organizing knowledge about failures, rehabilitation, and replacement alternatives of either vertical or battered earth-retaining walls. The problem is broken into two general models: retaining wall failure diagnosis and rehabilitation strategy design. An understanding of the domain knowledge and problem formulation is coupled with expert systems problem paradigm models and problem-solving strategies to create an intelligent aid for solving retaining wall problems.

BACKGROUND

Retaining Walls

Although numerous authors discuss the design of various types of new earth-retaining structures (*5–9*), little information is available on the rehabilitation of existing walls. Besides reliability and useful life expectancy, the factors that dictate the rational design of retaining walls are the available materials, magnitude, and type of forces acting on the wall, space required, soil characteristics, appearance, and, finally, cost. Yet, like most structures, retaining walls are designed largely from the designer's experience, which is often limited to a few types of walls. Although some might argue that the choice of design becomes fairly obvious once loads and geometric conditions have been established, one usually finds that retaining wall alternatives embrace a great variety of types, geometries, materials, and construction methods. As will be described, retaining walls can be divided into four main types—gravity, cantilever, anchored, and reinforced backfill—depending on how they retain earth and resist external loads.

Gravity Walls

Gravity walls provide stability against sliding and overturning by their weight alone or by their weight in combination with soil weight. Although such walls may be classified as either rigid or flexible, all are free to displace laterally, thereby mobilizing active earth pressures. Rigid gravity walls—which include cast-in-place mass concrete walls and masonry set block, brick, or rubble walls—have little tolerance to settlement. In addition, through-wall drainage is necessary to relieve hydrostatic pressure caused by saturated backfill soil. The components of flexible walls such as crib, bin, gabion, drystone rubble, and interlocking modular walls are not rigidly connected; this gives them the inherent advantages of positive drainage through the wall and some tolerance to nonuniform settlement. Crib walls are built by forming interconnected boxes from timber or concrete stringer-and-tie members. Examples of precast-concrete crib wall systems are Dura-Crib™, Dura-Hold™, Concrib™, and Criblock™. Bin walls are made of flexible sheeting such as corrugated aluminum or galvanized steel; examples include Terra-Wall™ and Armco™. Gabion walls are constructed from building blocks of metallic wire baskets filled with broken stone. Examples of interlocking precast-concrete units that may be dry set to form flexible gravity walls are PISA II™, Doublewal™, and Evergreen™. With some modular wall systems, the components themselves are backfilled as the wall is constructed, to provide added wall mass to resist lateral pressure. Typically the manufacturers of modular walls provide some guidelines for the design and construction of their system.

Cantilever Walls

Cantilever walls resist lateral forces through cantilever bending of the wall stem. Such walls include reinforced-concrete, soldier piles and lagging, and permanently anchored cylinder pile walls. Designing a conventional reinforced-concrete cantilever retaining wall involves determining the base size and the stem position on the base, as well as the thickness and reinforcing of the stem, toe, and heel. Design charts (*10*) based on the strength design method simplify the iterative process of designing concrete cantilever retaining walls. A minimum cost design method (*11*) may be used to proportion a cantilever retaining wall, satisfy design criteria, and minimize cost of concrete and steel. Design aids (*12*) are available for determining how bearing pressure varies with changes in footing width. To avoid diagonal tension cracks, the best corner reinforcing details depend on the toe-length-to-stem-thickness ratio of the wall (*13*). Soldier piles and timber lagging walls (also called Berlin walls) are constructed by first driving wide flange steel sections 6 to 10 ft apart along the wall line. As excavation in front of the wall proceeds, timber lagging is inserted between the piles. Soldier pile walls are not suitable for retaining soft clays or running soils such as loose sand, because portions of the wall must be left unsupported while lagging is installed. A permanently anchored cylinder pile wall (*14*) consists of prestressing steel strands in concrete cylinder piles anchored in rock. The steel tendons are bowed to pull the top of the caisson back toward the soil and thus reduce wall deflections.

Anchored Walls

Anchored walls support horizontal loads with tension members that are connected to the wall facing. These walls are constructed from the top down; this is an advantage because the slope is always supported and no preexcavation is required. Moreover, the wall's capacity may be increased, after the wall is completed, by installing additional anchors. One type of tieback wall (*14*) is an extension of the soldier pile and lagging wall, in that a row of high-strength steel tiebacks is placed into the soil behind the wall after each 5- to 10-ft lift is excavated. A curtain wall (*8*) consists of a "curtain" of reinforced concrete about 8 to 12 in. thick constructed from the top down as the slope is cut in benches. The wall is tied back with anchors at each bench level. There is no universal type of tieback anchor. Various soil strata require different anchorage systems and installation equipment: rock anchors if competent rock is near; augered earth anchors in cohesive soils; and small-diameter cased earth anchors in predominantly granular soil. Because tiebacks are soil dependent, if soil conditions are not determined accurately, problems such as pullout or inadequate development of frictional resistance may occur (*15*). Deadman anchors are short blocks, continuous beams, or plates that are placed in the excavated embankment behind the wall and then linked to face panels by reinforcing bars. The wall is then backfilled and compacted after the deadman anchors are positioned and connected. Deadman anchors derive their resistance from passive earth pressure. They can be implanted behind the wall near the ground surface, at depth, or in layers. Geotech™ is an example of a deadman retaining wall system. In the past, highway agencies have avoided building anchored walls out of concern that corrosion of the anchor tendons may lead to failures. Cheney (*16*) combined design criteria from several geotechnical contractors in a manual that contains basic anchoring concepts and guidelines for evaluating the feasibility of anchors in specific situations.

Reinforced Backfill Walls

Reinforced backfill walls include Reinforced Earth™, Retained Earth™, Georgia Stabilized Earth, Mechanically Stabilized Embankments™, Hilfiker Welded Wire Walls, and walls constructed with geotextiles or geogrids that provide tensile reinforcement within the backfill material. Reinforced backfill is a construction material composed of soil fill strengthened and mechanically stabilized by reinforcing that interacts frictionally with the soil. Reinforced backfill walls are constructed of horizontal reinforcement vertically spaced between layers of compacted backfill (*17–19*). The beneficial effects depend on a combination of the reinforcement's tensile strength and shear bond with the soil backfill. The reinforcing acts to supplement the tensile strength of the backfill, much the same as the steel bars in reinforced concrete. Various forms of reinforcing are used, such as steel strips, wire mesh strips, wire mesh fabric, geotextile fabric, and polymer grid. Reinforced backfill walls impose nearly uniform pressure distributions on foundation soils, thus creating uniform settlement. Reinforced backfill walls are a good choice if deep, poor foundation material exists at the site. In steel-reinforced systems, the reinforcing itself must be galvanized or epoxy coated to protect from corrosion (*20, 21*). The design assumption for these walls is that the reinforced material acts as a block, and the block acts as a gravity wall.

Expert Systems in Geotechnical and Transportation Engineering

Two reports (*3, 4*) summarize the current state of KBES in geotechnical and transportation engineering. Most of the systems discussed in this section are described in those reports.

WADI (*22*) and RETWALL (*23*) are two prototype systems that address aspects of retaining wall design that are closely related to the rehabilitation problem. WADI diagnoses concrete retaining wall failures in four categories: footing, drainage, or construction problems and problems associated with weak bearing soil. RETWALL is a selection system containing heuristics for selecting the appropriate types of retaining structure from a number of options. Neither of these systems is complete for practical application.

Expert system aids are of particular interest to transportation engineers, who generally do not have the time, training, or experience to handle maintenance and rehabilitation problems. Considerable work has been undertaken on highway pavement maintenance and rehabilitation problems (*24*). Pavement problems are the focus of systems such as ROSE (*25*) and PRESERVER (*26*). PARADIGM (*27*) is an integrated set of expert systems for the analysis and design of pavement rehabilitation strategies. Two component systems of PARADIGM and SCEPTRE and OVERDRIVE. SCEPTRE evaluates pavement surface conditions to identify workable rehabilitation and maintenance strategies. OVERDRIVE provides design guidance for asphalt concrete overlay rehabilitation strategies. Another component of PARADIGM optimizes a rehabilitation planning network according to construction budget constraints.

Several systems in other domains include ideas and modules that can be applied to the retaining wall rehabilitation problem. PROSPECTOR (*28*) is a mineral exploration interpretation system designed to handle problems in ore deposit identification, regional resource evaluation, and drilling site selection. In PROSPECTOR, ore deposit models are encoded as inference networks of connections or relations between field evidence and geological hypothesis. The term "model" refers to the body of knowledge about a particular ore deposit that is encoded into the system. Each ore deposit model is designed to emulate the reasoning process of an experienced exploration geologist who is assessing a given prospect site or region for the likelihood that it contains a deposit of the ore type represented by the model. In PROSPECTOR, both evidence and hypotheses are referred to as assertions, and models are referred to as inference networks of assertions. PROSPECTOR accommodates uncertainty in both evidence and hypotheses using an approximate form of Bayesian probability. Inference networks of assertions along with Bayesian decision theory have been used to develop a consultant system for physical property prediction of chemicals (*29*). The same approach is described here to diagnose failing retaining walls.

OVERVIEW OF RETAIN

The expert system RETAIN, described herein, is designed to function as an aid to designers whose personal experience with failure diagnosis, retaining wall design, and cost estimation is limited to a narrow range of circumstances and design alternatives. Certainly, this expert system is not to be viewed as a substitute for an experienced knowledgeable practitioner. Instead, it represents an attempt to frame the stated problem so that gaps in knowledge and experience can be identified but permits its continual evolution as practitioners contribute new knowledge. Although the inevitable limitations of expert systems often are a source of frustration to users, the potential of these systems to identify critical analysis and design alternatives should not be underestimated. Familiarity with these alternatives is most often the weakest component in the repertoire of even the most experienced individual.

The functional overview of RETAIN is shown in Figure 1. The system consists of a series of modules: Site Identification, Failure Diagnosis, Strategy Synthesis, Strategy Analysis, Cost Estimation, and Evaluate/Consistency; each completes a subtask of the solution as a whole. The modules are executed by the user from a menu in the main controlling program. The system requires a data base for storing knowledge such as soil properties, construction material properties, and cost estimation data as well as problem descriptions and intermediate results.

The Site Identification module prepares a geometric model of the site including the wall, soil, loading, and surroundings. The wall parameters are type, location, construction material, and dimensions. The module contains knowledge that helps assign soil properties by soil characteristics. Surcharge loads are modeled using four primitive load types: point, strip, uniform, and line. For example, a highway running parallel to the wall line may be modeled as a strip load of uniform magnitude. The class of roadway loads may be further subdivided according to number of lanes and usage. The appropriate width and magnitude of various roadway types may be collected in a data base along with knowledge about other surcharge load types. Surrounding site conditions include soil slopes and clear distances in front of and behind the wall. The function of the site identification module is to assemble problem details required by the other modules.

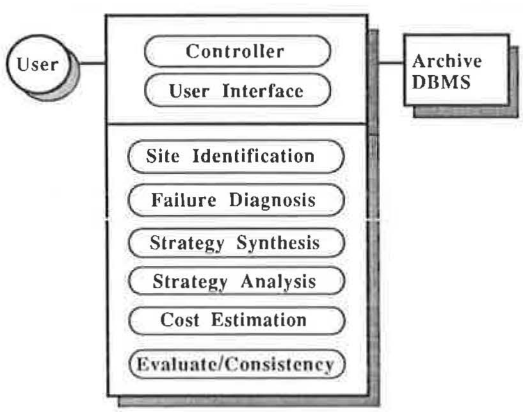

FIGURE 1 Functional overview of RETAIN.

Expert-system problem solving techniques tend to follow either the derivation or the formation approach (*4*). In the RETAIN system, retaining wall failure diagnosis follows the derivation approach, whereas rehabilitation strategy synthesis follows the formation approach. In the derivation approach, the most appropriate solution is selected from a set of predefined solutions stored in the knowledge base. To implement the derivation approach, all possible solutions to a problem must be listed, and each solution must be potentially justified by the definition of any given problem. A knowledge base for solving derivation problems is usually represented by an inference network with connections between the predefined solutions and the input data. The derivation approach may be useful for solving diagnosis, interpretation, and classification problems. On the other hand, the formation approach is preferred when all possible solutions to a problem (for practical reasons) cannot or should not be listed and stored in the knowledge base. The formation approach is implemented by forming complete solutions from components of solutions stored in the knowledge base. The formation approach requires that solution components and the heuristics for combining the components be predefined. Formation problems are generally planning and design problems.

In general, diagnosis means finding the state of a system by interpretation of relevant data. Diagnosing a failing retaining wall involves evaluating the stability and failure characteristics of the wall to determine the influencing mechanisms. There are two ordered steps in the Failure Diagnosis module: retaining wall stability analysis using conventional (algorithmic) programming, followed by diagnostic inferencing. Retaining wall analysis means the application of engineering knowledge, including equations and algorithms for computing earth and surcharge pressures. The pressures are combined to determine resultant wall loads. The overall stability, considering bearing capacity, overturning, and sliding, and the structural stability, including bending in wall stems or tension in tiebacks, are computed. Although analysis is implemented by equations and algorithms, it is not independent of wall type. For instance, computing the bending capacities of various cantilever wall types, such as soldier piles and lagging or cast-in-place reinforced concrete, follows different algorithms. The diagnostic knowledge base is represented by inference networks similar to those in the PROSPECTOR system (*28*), with logical and plausible relations to reason with uncertainty. Diagnostic inferencing is handled by DIGR (*30*), an OPS diagnostic inference engine. The DIGR kernel and an example of a retaining wall diagnostic network model are discussed in the following section.

The synthesis, analysis, and evaluation of rehabilitation strategies constitute an engineering design problem. Solving such a problem involves a number of distinctive phases, beginning with the definition of a particular problem and ending with the selection of an optimum solution. Commonly the three main phases are as follows:

1. Synthesis of preliminary design. The conceptual aspects of engineering design are embodied in this phase. The result of synthesis is a set of preliminary design alternatives that satisfies the end product specifications. The specifications for retaining wall rehabilitation designs are the wall type to be rehabilitated, failure mechanisms to be corrected, limitations of the soil, and constraints dictated by the site. Synthesis deals with the formation of design alternatives by searching and checking subsystems. In the RETAIN system, this task is accomplished by the Strategy Synthesis module.
2. Analysis. During this phase preliminary design alternatives are studied using mathematical and other analysis procedures. The purpose of this phase is to refine the preliminary designs in an effort to determine their response in the intended environment. Important aspects of this phase are selection of the proper analysis procedure, correct use of these procedures, and appropriate interpretation of the results. The Strategy Analysis module sizes the components of each synthesized rehabilitation strategy. Preliminary designs include wall and footing dimensions, drainage characteristics, approximate spacing of tiebacks and anchors, amount of excavation, and so on.
3. Evaluation and optimization. This phase, composed of the Cost Estimation and Evaluate/Consistency modules, involves evaluation of the analyzed design alternatives. The Cost Estimate module prepares a preliminary cost estimate for each design. Cost estimation may be handled by production rules, with algorithmic computational functions, which query a data base for specific unit cost factors (*31*). In the Evaluate/Consistency module, all feasible rehabilitation strategies are ranked by how well they meet project objectives such as cost, availability of materials, minimizing construction time, and minimizing disruptions to traffic and other structures. Output from this phase is the overall evaluation and ranking of feasible rehabilitation alternatives.

The RETAIN system is represented by rules in an OPS language (*32, 33*) production system. A rule-based production system architecture was chosen because it is widely available and easily transported. The OPS languages provide flexibility for storing domain-specific knowledge and for alternative inference engine strategies for problem solving.

The knowledge base of the expert system is being obtained from three sources. First, literature sources, which include technical journals, textbooks, manuals, public and commercial documents and reports, are used. Although the proliferation of retaining wall alternatives currently available has created a greatly enlarged set of design options (*1*), the issues pertaining to retaining wall rehabilitation are rarely addressed in the literature, except for a few isolated case studies. Useful published sources are case studies and technical documents that describe infrastructure rehabilitation techniques such as underpinning and anchoring (*16, 34*). Knowledge obtained from literature sources is being used in soil property assignment, wall stability analysis, and rehabilitation strategy analysis. A second source of domain-specific knowledge is a panel of experts who will help develop retaining wall diagnostic networks and provide information such as applicable heuristics for cost estimation and conditions under which particular rehabilitation techniques are feasible. Knowledge from the third source, testing, will be incorporated during the final stages of development when a number of case tests are undertaken for validation and further development. Results of testing will be incorporated into the system as appropriate.

USE OF INFERENCE NETWORKS TO DIAGNOSE RETAINING WALL FAILURES

In this section, an example inference network for retaining wall failure diagnosis is described. The network is a form of the inference network of assertions used by the PROSPECTOR (*28*) system to identify ore deposits. The DIGR (*30*) inference engine is used to traverse the network and apply logical and plausible relations so it can reason with uncertainty.

Network Model

An inference network can be depicted in hierarchical form, as shown in Figure 2, in which the terminal or leaf nodes correspond to evidence and the intermediate nodes represent conclusions or hypotheses. All nodes are referred to as assertions. The certainty of the top assertion (Forward Tilting Failure) indicates how well the available evidence matches the network model. The certainty of the top assertion is determined by the second-level assertions (Forward Tilting Wall and Forward Tilting Failure Type), which in turn are determined by the third-level assertions, and so forth. The search is finished when assertions are reached that may be established from evidence assertions. This tree traversal, executed as backward chaining, is depth-first with a left-to-right sequence.

To illustrate inference network modeling, first consider the failure modes shown in Figure 3. These failures are common for cast-in-place concrete walls, such as reinforced-concrete cantilever, buttress, and counterfort walls, and mass gravity walls. The primary evidence for each of the failure types is forward tilting of the wall. However,

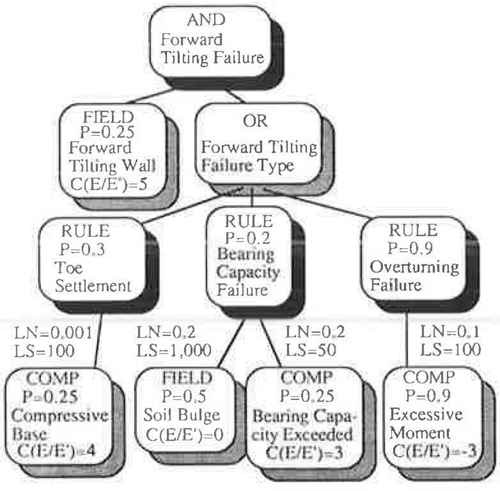

FIGURE 2 Example of inference network of assertions for wall diagnosis.

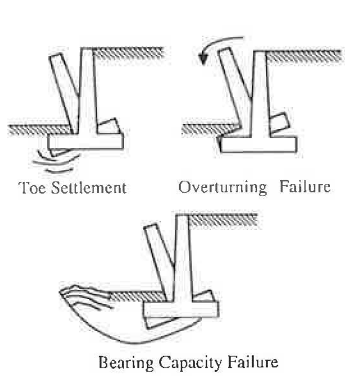

FIGURE 3 Forward tilting failures of cast-in-place concrete walls.

the cause of tilting is different in each case. Toe settlement may have occurred if the bearing soil is compressive. Compressive soils include soft clays and loose sands or gravels. Overturning failure occurs when the earth pressure and surcharge pressures cause an overturning moment about the wall toe that is greater than the resisting moment. If the foundation soil is completely rigid, the wall will rotate about its toe. Bearing capacity failure occurs when the bearing pressure is higher than the bearing capacity of the foundation soil. A bearing capacity failure causes foundation soil to be pushed from beneath the wall toe, causing a bulge in front of the wall.

The inference network in Figure 2 models the failure modes shown in Figure 3. For given evidence, the network is used to assess the likelihood of each failure mode. There are two types of evidence nodes: COMP and FIELD. COMP nodes represent evidence that may be gathered from the problem description or from retaining wall stability analysis. FIELD nodes represent evidence supplied by the user after field inspection. There are three types of hypothesis node: AND, OR, and RULE. AND and OR nodes use logical relations to combine subnode evidence, whereas RULE nodes use plausible relations.

Reasoning with Uncertainty

When a retaining wall is diagnosed, some uncertainty exists whether evidence and hypotheses are true or false. The certainty measure $C(E/E')$ of an evidence node may be described as the certainty of the node evidence E given the available evidence E'. Certainty measures are defined on a scale of 5 (certainly true) to −5 (certainly false). Initially the certainty of each node in a network is unknown (certainty = 0). As evidence is gathered by using logical and plausible relations some hypotheses may be definitely established, whereas others may become only more or less likely or even excluded entirely. Verbal descriptions of numerical certainties are given in Table 1.

In the retaining wall diagnosis network, a set of evidence certainty measures is assigned by the user and the system for each wall diagnosed. Evidence certainty measures are provided by the user at FIELD nodes. For example, if the field evidence shows that a wall is tilting forward, the input certainty at the FIELD node Forward Tilting Wall is 5 (see Figure 2). In some cases the user may not have enough information to determine if evidence is present or absent. An example of this situation might occur at node Soil Bulge. Because of soil erosion, vegetation growth, or construction activity in front of the wall, the user may not be able to determine if the Soil Bulge evidence is present or absent. In this case, the user may indicate that there is "no information," and the evidence certainty measure remains equal to zero (see Figure 2). At COMP nodes, evidence certainty measures are assigned by the system from the problem description. For example, the evidence certainty at node Compressive Base need not be provided by the user, because it may be inferred from the wall's bearing soil properties. Similarly, the certainty at node Excessive Moment can be determined by comparing the calculated overturning moments with resisting moments. Knowledge for assigning certainty measures at COMP nodes is provided by domain experts.

TABLE 1 VERBAL DESCRIPTION OF CERTAINTY MEASURES

Verbal Description	Certainty Measure $C(E/E')$
Certain	5
Very likely	4
Likely	3
Maybe	2
Slightly likely	1
No information	0
Slightly unlikely	−1
Maybe not	−2
Unlikely	−3
Very unlikely	−4
Certainly not	−5

Logical Relations

The fuzzy-set formulas of Zadeh (*35*) are used to evaluate logical relations. Two primitive operations, conjunction (AND) and disjunction (OR), constitute the logical relations. If several assertions leading to a hypothesis must be true for the hypothesis to be true, the hypothesis is the conjunction of the assertions. The probability of a logical conjunction is the minimum probability of its subnodes. A disjunction of several assertions occurs when only one of the assertions must be true for the hypothesis to be true. A logical disjunction is assigned a probability value equal to the maximum probability of its subnodes. In the example network, the probability that the available evidence matches a Forward Tilting Failure Type represents a disjunction equal to the maximum probability that the evidence matches any of the forward tilting models Toe Settlement, Bearing Capacity Failure, or Overturning Failure. Furthermore, the probability of the conjunction node Forward Tilting Failure is the minimum probability of a Forward Tilting Wall and a Forward Tilting Failure Type, because the probability of a Forward Tilting Failure Type is not needed when there is no Forward Tilting Wall.

One problem with logical relations is that in the case of conjunction, if all but one of the assertions can be established the probability often remains unchanged from the case when none of the assertions was known. This effect may be desirable, such as at the node Forward Tilting Wall, but to allow some "partial credit" for uncertain evidence, plausible relations are used.

Plausible Relations

A RULE node represents domain knowledge that is to be confirmed or refuted by its subnodes. The probability of RULE nodes is governed by plausible relations. Plausible relations (*28, 29, 35*) are based on Bayesian decision theory, assuming the evidence is conditionally independent. Plausible relations are simply expressed in terms of odds O rather than probability P, where

$$P = O/(1 + O) \tag{1}$$

or

$$O = P/(1 - P) \tag{2}$$

The "odds-likelihood" form of Bayes's rule is

$$O(H/E) = LS \times O(H) \tag{3}$$

when evidence is present, or

$$O(H/-E) = LN \times O(H) \tag{4}$$

when evidence is absent.

$O(H/E)$ and $O(H/-E)$ are the posterior odds on the hypothesis H given that evidence E is either present or absent and $O(H)$ is the prior odds on the hypothesis. LS and LN are nonnegative likelihood ratios. Using the relationships expressed by Equations 1 and 2, we may write Equations 3 and 4 in terms of probability P:

$$P(H/E) = \frac{LS\ P(H)}{(LS - 1)P(H) + 1} \tag{5}$$

$$P(H/-E) = \frac{LN\ P(H)}{(LN - 1)P(H) + 1} \tag{6}$$

Likelihood ratios are associated with each plausible relation. A likelihood ratio measures the degree to which a change in the probability of the evidence assertion changes the probability of the hypothesis. The sufficiency measure LS is the degree of support of a hypothesis given positive evidence, whereas the necessity measure LN is the degree of refutation when evidence is absent, that is,

If Evidence (E),
Then [to degree LS, LN] Hypothesis (H)

In contrast to certainty measures, which must be provided for each wall diagnosed, likelihood ratios are fixed in each diagnosis model. Initially, when the model is created, they are assigned by a domain expert (although when the network is tested, one or more ratios may be changed to improve diagnostic performance). Verbal descriptions of the range of numerical likelihood ratios are given in Table 2.

When there are n conditionally independent subnodes contributing to the certainty of a RULE node, the plausible relation that combines all evidence is

$$O(H/E') = [L_1' \ L_2' \ L_3' \ \ldots \ L_n'] \ O(H) \tag{7}$$

where L_i' is the effective likelihood ratio of the ith piece of evidence E_i.

$$L_i' = \frac{O(H/E_i')}{O(H)} = \frac{P(H/E_i')[1 - P(H)]}{[1 - P(H/E_i')]P(H)} \tag{8}$$

TABLE 2 VERBAL DESCRIPTION OF LIKELIHOOD RATIOS

Verbal Description	Likelihood Ratio
Completely sufficient	1,000,000
Extremely sufficient	10,000
Very suggestive	100
Moderately suggestive	10
Mildly suggestive	5
Weakly suggestive	2
Indifferent	1
Weakly necessary	0.5
Mildly necessary	0.2
Moderately necessary	0.1
Very necessary	0.01
Extremely necessary	0.0001
Completely necessary	0.000001

Prior probability of truth is required for the evidence $P(E)$ and hypothesis $P(H)$ node of plausible relations. The prior probabilities are assigned by domain experts during the construction of the network mode. Verbal descriptions of prior probabilities listed in Table 3 may be useful.

Because the user may not be able to state that E is definitely present or absent, he responds in terms of certainty, $C(E/E')$, where E' denotes the observation that causes him to suspect the presence of E. The user's response is converted to posterior probability $P(E/E')$ so that $P(E/E') > P(E)$ when evidence is present absent $[C(E/E') > 0]$. Using the certainty measures has absent $[C(E/E') > 0]$. Using the certainty measures has the advantage that it takes into account the prior probability of evidence. Thus the user need not know $P(E)$ as a reference point when expressing the presence or absence of the evidence.

Posterior probability is mapped as a piecewise linear function of certainty normalized with respect to the prior probability, so that for certainty scaled between 5 and −5,

$$P(E/E') = 1 \quad \text{when } C(E/E') = 5$$

$$P(E/E') = P(E) \quad \text{when } C(E/E') = 0$$

$$P(E/E') = 0 \quad \text{when } C(E/E') = -5$$

Thus the following function is used to compute the posterior probability of each piece of evidence:

$$P(E/E') = \frac{C(E/E')[1 - P(E)]}{5} + P(E) \quad \text{when } C(E/E') > 0 \qquad (9)$$

$$P(E/E') = \frac{C(E/E')P(E)}{5} + P(E) \quad \text{when } C(E/E') < 0 \qquad (10)$$

A piecewise linear function of $P(E/E')$ is used to update the conditional probability $P(H/E')$, such that

$$P(E/E') = 0 \quad \text{when } P(H/E') = P(H/-E)$$

$$P(E/E') = P(E) \quad \text{when } P(H/E') = P(H)$$

$$P(E/E') = 1 \quad \text{when } P(H/E') = P(H/E)$$

TABLE 3 VERBAL DESCRIPTION OF PRIOR PROBABILITIES *P*

Verbal Description	Prior Probability (*P*)
Always present	0.999
Almost always present	0.99
Abundant	0.9
Very common	0.75
Common	0.5
Fairly common	0.25
Occasional	0.1
Rare	0.01
Extremely rare	0.001

The updating equation is then

$$P(H/E') = P(H) + \frac{P(H/E) - P(H)}{1 - P(E)} \times [P(E/E') - P(E)] \quad \text{when } P(E/E') > P(E) \qquad (11)$$

$$P(H/E') = P(H/-E) + \frac{P(H) - P(H/-E)}{P(E)} \times P(E/E') \quad \text{when } P(E/E') < P(E) \qquad (12)$$

Conditional probability may be converted to the conditional certainty using Equations 9 and 10, or

$$C(H/E') = 5\,\frac{P(H/E') - P(H)}{1 - P(H)} \quad \text{when } P(H/E') > P(H) \qquad (13)$$

$$C(H/E') = 5\,\frac{P(H/E') - P(H)}{P(H)} \quad \text{when } P(H/E') < P(H) \qquad (14)$$

Example Diagnosis

To see how inferencing works, consider the network in Figure 2. The top node conjunction, Forward Tilting Failure, assumes the lowest probability of its subnodes, Forward Tilting Wall and Forward Tilting Failure Type. The top node probability cannot be assigned until the logical disjunction at the Forward Tilting Failure Type node is assigned. However the disjunction cannot be determined until the probability of its subnode plausible relations, Toe Settlement, Bearing Capacity Failure, and Overturning Failure, are evaluated.

Suppose the input certainties for the leaf nodes are as follows:

$$C(E/E')_{\text{Forward Tilting Wall}} = 5$$
$$C(E/E')_{\text{Compressive Base}} = 4$$
$$C(E/E')_{\text{Soil Bulge}} = 0$$
$$C(E/E')_{\text{Bearing Capacity Exceeded}} = 3$$
$$C(E/E')_{\text{Excessive Moment}} = -3$$

From the network model we have the prior probabilities of the evidence nodes,

$$P(E)_{\text{Forward Tilting Wall}} = 0.25$$
$$P(E)_{\text{Compressive Base}} = 0.25$$
$$P(E)_{\text{Soil Bulge}} = 0.5$$
$$P(E)_{\text{Bearing Capacity Exceeded}} = 0.25$$
$$P(E)_{\text{Excessive Moment}} = 0.9$$

We can now compute the posterior probability of the evidence. From Equation 9,

$$P(E/E')_{\text{Forward Tilting Wall}} = \frac{5(1 - 0.25)}{5} + 0.25 = 1$$

$$P(E/E')_{\text{Compressive Base}} = \frac{4(1-0.25)}{5} + 0.25 = 0.85$$

$$P(E/E')_{\text{Soil Bulge}} = \frac{0(1-0.5)}{5} + 0.5 = 0.5$$

$$P(E/E')_{\text{Bearing Capacity Exceeded}} = \frac{3(1-0.25)}{5} + 0.25 = 0.7$$

From Equation 10,

$$P(E/E')_{\text{Excessive Moment}} = \frac{-3 \times 0.9}{5} + 0.9 = 0.36$$

The probability of a particular hypothesis given the evidence $P(H/E)$ is computed directly for nodes Toe Settlement and Overturning Failure, each of which has one evidence subnode. Components of $P(H/E)$ are computed for node Bearing Capacity Failure, which has more than one subnode. Because $P(E/E') > P(E)$ at nodes Compressive Base, Soil Bulge, and Bearing Capacity Exceeded, the sufficiency ratio and Equation 5 are used to compute $P(H/E)$. Thus

$$P(H/E)_{\text{Toe Settlement}} = \frac{100 \times 0.3}{(100-1) \times 0.3 + 1} = 0.977$$

and for each evidence node of Bearing Capacity Exceeded,

$$P(H/E)_{\text{Soil Bulge}} = \frac{1000 \times 0.2}{(1000-1) \times 0.2 + 1} = 0.996$$

$$P(H/E)_{\text{Bearing Capacity Exceeded}} = \frac{50 \times 0.2}{(50-1) \times 0.2 + 1} = 0.977$$

The necessity ratio and Equation 6 are used to compute $P(H/E)$ at node Excessive Moment because $P(E/E') < P(E)$.

$$P(H/-E)_{\text{Overturning Failure}} = \frac{0.1 \times 0.9}{(0.1-1) \times 0.9 + 1} = 0.474$$

$P(H/E)$ is updated to $P(H/E')$ using Equation 11 or 12.

$$P(H/E')_{\text{Toe Settlement}} = 0.3 + \frac{0.977 - 0.3}{1 - 0.25} \times (0.85 - 0.25) = 0.842$$

$$P(H/E')_{\text{Soil Bulge}} = 0.2 + \frac{0.996 - 0.2}{1 - 0.5}(0.5 - 0.5) = 0.2$$

$$P(H/E')_{\text{Bearing Capacity Exceeded}} = 0.2 + \frac{0.926 - 0.2}{1 - 0.25} \times (0.7 - 0.25) = 0.636$$

$$P(H/E')_{\text{Overturning Failure}} = 0.474 + \frac{0.9 - 0.474}{0.9} \times 0.36 = 0.644$$

The plausible relation for combining evidence (Equation 7) is used to determine the odds of Bearing Capacity Failure, which has more than one piece of evidence. First, effective likelihood ratios for the evidence are found from Equation 8. Hence,

$$L'_{\text{Soil Bulge}} = \frac{0.2(1-0.2)}{(1-0.2)0.2} = 1$$

$$L'_{\text{Bearing Capacity Exceeded}} = \frac{0.636(1-0.2)}{(1-0.636)0.2} = 6.98$$

Then, using Equation 7,

$$O(H/E')_{\text{Bearing Capacity Failure}} = (1 \times 6.98) \times 0.25 = 1.74$$

where $O(H) = 0.2/(1 - 0.2) = 0.25$ from Equation 2. Then from Equation 1,

$$P(H/E')_{\text{Bearing Capacity Failure}} = \frac{1.74}{1 + 1.74} = 0.636$$

This example illustrates the potential advantage of plausible relations over logical relations. Notice that $P(H/E')$ at node Bearing Capacity Failure is equal to $P(H/E')$ of its subnode Bearing Capacity Exceeded. This is because the plausible relation allows the other subnode, Soil Bulge, with $C(E/E') = 0$ ("no information") to contribute "nothing" to the conditional probability of the hypothesis node Bearing Capacity Failure. Had the node Bearing Capacity Failure been a logical conjunction (AND node), its probability would have been assigned the lowest value of its subnodes Soil Bulge and Bearing Capacity Exceeded. Thus $P(H/E')$ at node Bearing Capacity Failure would have been assigned a value of 0.2 corresponding to the node Soil Bulge, and the evidence at node Bearing Capacity Exceeded would have been ignored.

We have seen how plausible relations are implemented. We now have the posterior probability of each subnode of the disjunction, Forward Tilting Failure Type, which we may convert to certainty using Equation 13 or 14.

$P(H/E')_{\text{Toe Settlement}} = 0.842;$

$$C(H/E')_{\text{Toe Settlement}} = 5\left(\frac{0.842 - 0.3}{1 - 0.3}\right) = 3.87$$

$P(H/E')_{\text{Bearing Capacity Failure}} = 0.636;$

$$C(H/E')_{\text{Bearing Capacity Failure}} = 5\left(\frac{0.636 - 0.2}{1 - 0.2}\right) = 2.73$$

$P(H/E')_{\text{Overturning Failure}} = 0.644;$

$$C(H/E')_{\text{Overturning Failure}} = 5\left(\frac{0.644 - 0.9}{0.9}\right) = -1.42$$

From the three nodes that are connected through the disjunction at node Forward Tilting Failure Type, the one with the maximum certainty $P(H/E')$ is selected, in this case Toe Settlement, and assigned that probability value as well as its prior probability to the node Forward Tilting Failure, that is,

$$P(H/E')_{\text{Forward Tilting Failure Type}} = \max\{0.842, 0.636, 0.644\} = 0.842$$

$$P(H)_{\text{Forward Tilting Failure Type}} = 0.3$$

$$C(H/E')_{\text{Forward Tilting Failure Type}} = 5\left(\frac{0.842 - 0.3}{1 - 0.3}\right) = 3.87$$

The top node conjunction, Forward Tilting Failure, assumes the lowest conditional probability of its subnodes, Forward Tilting Wall and Forward Tilting Failure Type, and the prior probability of the subnode with the lowest conditional probability, that is,

$$P(H/E')_{\text{Forward Tilting Failure}} = \min\{1, 0.842\} = 0.842$$

$$P(H)_{\text{Forward Tilting Failure}} = 0.3$$

$$C(H/E')_{\text{Forward Tilting Failure}} = 3.87$$

In summary, for the given evidence the certainty measures shown in Table 4, on a scale of −5 to 5, were found for each node in the network model.

The certainty of the top node, Forward Tilting Failure, indicates how well the available evidence matches the network model. A certainty measure of 3.87 suggests that it is "very likely" that our evidence matches the model. The certainty at nodes Toe Settlement, Bearing Capacity Failure, and Overturning Failure indicates the certainty that each of these failure modes is responsible for causing our wall to tilt forward. From the evidence, it is "very likely" ($C = 3.87$) that failure is caused by Toe Settlement, "likely" ($C = 2.73$) that we have a Bearing Capacity Failure, and "maybe not" or "slightly unlikely" that the failure is caused by an Overturning Failure. Notice that the certainty of Toe Settlement is greater than that of Bearing Capacity Failure. This result is reasonable because if a cast-in-place concrete wall experiences bearing capacity failure, toe settlement has already occurred.

Once a network model is built, test cases are applied to validate the model. Results of the test cases may indicate that the model is incorrect. To improve the performance of the model, prior probabilities or likelihood ratios, or both, may be modified. Inferencing can be improved also by adding new nodes and links.

TABLE 4 CERTAINTY MEASURE BY NODE

Node	Certainty Measure
Forward Tilting Failure	3.87
Forward Tilting Wall	5.00
Forward Tilting Failure Type	3.87
Toe Settlement	3.87
Bearing Capacity Failure	2.73
Overturning Failure	−1.42
Compressive Base	4.00
Soil Bulge	0.00
Bearing Capacity Exceeded	3.00
Excessive Moment	−3.00

Diagnostic Inference Engine

The RETAIN prototype retaining wall diagnosis networks are processed using the DIGR (*30*) inference engine. DIGR is based on SRI's PROSPECTOR system and its RENE acquisition shell. The portion of DIGR that interprets diagnostic networks is an OPS5 (*34*) production system. DIGR recognizes a variety of network node types including the COMP, FIELD, RULE, AND, and OR nodes used in the example network. DIGR has some supporting user and rule tracing explanation facilities. The system prompts for input on certainty measures associated with FIELD nodes. DIGR summarizes intermediate and final results of a diagnostic session. During a session, DIGR allows the user to revise certainty measures assigned to FIELD nodes.

DIGR may be used to traverse any diagnostic network that falls into a pattern of organization, and may have some nodes serve as hypotheses ("probable causes") and others serve as evidence ("symptoms"). The logic of how symptoms combine to support probable causes determines how the various symptoms are queried by the user. Symptom queries are executed in a depth-first, left-to-right sequence. The order in which symptoms are queried can be controlled by positioning nodes near other contributing symptom nodes. The certainty of a hypothesis node is updated after the certainty of each of its subnodes is determined. DIGR produces a listing of final certainties of all nodes for an observed set of faults or abnormal behavior represented by the leaf nodes.

STRATEGY SYNTHESIS

Formation problems such as the synthesis of rehabilitation designs may be solved using a generate-and-test strategy (*4*), in which all possible solutions are generated from solution components in the knowledge base and tested until a solution is found that satisfies the goal specifications. The plan-generate-test strategy is a version of generate-and-test that restricts the number of possible solutions by pruning inconsistent solutions. Pruning is achieved by a planning stage at which data are interpreted and constraints are evaluated; these constraints eliminate solution components and component combinations that are inconsistent.

The plan-generate-test strategy is implemented in the RETAIN system to form complete rehabilitation designs from design components generated for the failure mechanism that is to be corrected. Associated with each wall failure is knowledge regarding which strategy or strategies may be used for rehabilitation. For example, alternatives for a structurally sound gravity wall that is exhibiting excessive lateral movement might be one or more of the following:

- Remove existing wall and regrade embankment (in the absence of a new wall).
- Add an earthen berm to stabilize the existing wall.
- Relieve soil pressure, hydrostatic pressure, and surcharge on wall.
- Brace or tie wall with anchors.
- Replace wall with another gravity wall or some alternative (cantilever, soldier piles, tieback, etc.).

A hierarchical network of retaining wall rehabilitation strategies is shown in Figure 4. Three approaches taken when rehabilitating a failing retaining wall are: to repair the wall, to remove the wall (then regrade), or to build a new wall. Retaining wall repair methods are grouped into two categories, upgrade and maintain. Upgrade repairs are performed to increase the load carrying capacity of an existing wall or its foundation, whereas maintenance repairs improve the efficiency, life expectancy, or aesthetics of an existing wall. As shown in Figure 4, upgrade repairs are underpin, buttress, grout wall base, and tieback. Figure 4 shows the hierarchical decomposition of maintenance and upgrade repair strategies. For example, there are three ways to improve drainage. They are to clean existing drains, to install wall drains, or to install surface drains. As another example, retaining wall underpinning techniques include pile underpinning and pit or pier underpinning, and pile underpinning methods include root piles and bracket piles. The hierarchical network shows the decomposition of the set of all rehabilitation strategies into components. A strategy component can be uniquely identified by following the network from a terminal node to the top. For example, two strategy components are rock anchors tieback upgrade repair strategy and reinforced backfill replacement strategy.

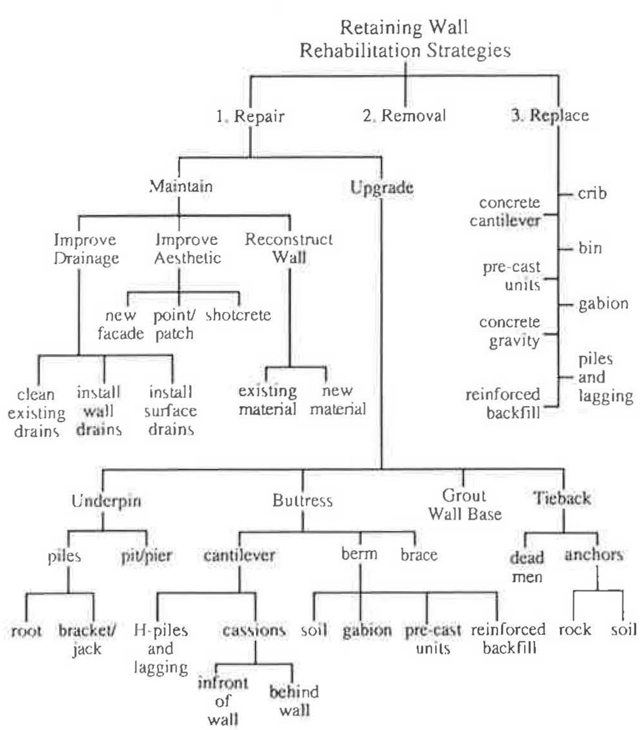

FIGURE 4 Hierarchical network of retaining wall rehabilitation strategies.

Associated with each rehabilitation strategy is a set of feasibility constraints. Constraints include wall type to be rehabilitated, limitations of the soil, and restrictions dictated by the site. The following list includes a selection of conditions under which replacement by soldier piles and lagging may be technically infeasible:

- Height of wall is less than 10 ft and length is less than 100 ft,
- Soil beneath the base of wall has insufficient strength (specified by the expert system for the given type and height of backfill),
- Access for pile driving or drilling equipment is inadequate,
- Operation of construction equipment creates an unacceptable disturbance (such as vibrations or diversion of traffic) to the environment,
- If uncorable rock exists at a depth, below the base of the wall, equal to one-half the height of the wall, or
- Subsidence-sensitive structures are located above the wall at a horizontal distance less than the height of the wall.

The heuristic knowledge of which strategies should be considered as possible candidates for rehabilitating a given wall failure may be represented by a relational data base table. The table represents the many-to-many relationship between two columns, one for failure mode and one for strategy component. The intersection and join operations of a data base management system data manipulation language may be used to generate the pruned set of solution components and complete solutions from the solution components. Two other tables necessary for synthesis are wall failures and associated certainties, and strategy components and associated feasibilities.

CONCLUSION

In this paper, the architecture for retaining wall rehabilitation design was described. This prototype system will be extended and field tested in 1988. Although the focus of this paper was RETAIN, the same development framework and application experience could be transferred to other infrastructure elements such as pipe networks or pavements.

ACKNOWLEDGMENT

Support for this research was provided by the National Science Foundation.

REFERENCES

1. K. A. Godfrey, Jr. Retaining Walls: Competition or Anarchy? *Civil Engineering*, ASCE, Vol. 54, Dec. 1984, pp. 48–52.
2. D. Donovan. Over the Hill; Aging Retaining Walls Next City Engineering "Crisis". *Pittsburgh Press*, April 13, 1986, p. 5.
3. S. S. Kim, M. L. Maher, R. E. Levitt, M. F. Rooney, T. J. Siller, and S. G. Richie. *Survey of the State-of-the-Art Expert/Knowledge Based Systems in Civil Engineering*. USA-CERL Special Report P-87/01. U.S. Army Corps of Engineers, Construction Engineering Research Laboratory, Champaign, Ill., Oct. 1986.
4. M. L. Maher, ed. *Expert Systems for Civil Engineers: Technology and Application*. ASCE, New York, 1987.
5. N. M. Newmark and W. J. Hall, eds. *Foundation Design*. Prentice-Hall, Inc., New York, 1962.
6. G. A. Leonards, ed. *Foundation Engineering*. Civil Engineering Series. McGraw Hill, New York, 1962.
7. K. Terzaghi and R. B. Peck. *Soil Mechanics in Engineering Practice*. John Wiley and Sons, New York, 1967.
8. R. E. Hunt. *Geotechnical Engineering Analysis and Evaluation*. McGraw-Hill, Inc., New York, 1986.
9. L. Casagrande. Comments on Conventional Design of Retaining Structures. *Journal of the Soil Mechanics and Foundations Division*, ASCE, Vol. 99, No. SM2, Feb. 1973, pp. 181–198.
10. M. M. Gupta and L. L. Friel. Design Aids for Cantilever Retaining Walls. *Journal of the Structural Division*, ASCE, Vol. 103, No. ST5, May 1977, pp. 1113–1126.
11. E. J. Rhomberg and W. M. Street. Optimal Design of Retaining Walls. *Journal of the Structural Division*, ASCE, Vol. 107, No. ST5, May 1981, pp. 992–1002.
12. J. W. Stiles. Practical Aspects of Cantilever Retaining Wall Design. In *Highway Research Record 302*, HRB, National Research Council, Washington, D.C., 1970, pp. 87–96.
13. I. H. E. Nilsson and A. Losberg. Reinforced Concrete Corners and Joints Subjected to Bending Moment. *Journal of the Structural Division*, ASCE, Vol. 102, No. ST6, June 1976, pp. 1229–1254.
14. P. J. Nicholson and D. L. Boley. Alternative Methods for Retaining Walls. *Public Works*, Vol. 116, No. 5, May 1985, pp. 66–68.
15. R. H. DiLouie, Jr. Practical Considerations in Tieback Construction. *Journal of the Construction Division*, ASCE, Vol. 107, No. CO2, June 1981, pp. 181–191.
16. R. S. Cheney. *Permanent Ground Anchors*. Technical Report FHWA-DP-68-1R. FHWA, U.S. Department of Transportation, Nov. 1984.
17. H. Vidal. The Principle of Reinforced Earth. In *Highway Research Record 282*, HRB, National Research Council, Washington, D.C., 1969, pp. 1–16.
18. K. L. Lee, B. D. Adams, and J.-M. J. Vagneron. Reinforced Earth Retaining Walls. *Journal of the Soil Mechanics and Foundations Division*, ASCE, Vol. 99, No. SM10, Oct. 1973, pp. 745–764.
19. T. S. Ingold. *Reinforced Earth*. Thomas Telford Ltd., London, 1982.
20. R. Y. Bush, and B. M. Ghadiali. Reinforced Earth Used at Palo Verde Nuclear Plant. *Journal of the Construction Division*, ASCE, Vol. 107, No. CO1, March 1981, pp. 121–126.
21. S. Frondistou-Yannas. *Corrosion Susceptibility of Internally Reinforced Soil Retaining Walls*. Research Report FHWA/RD-83/105. FHWA, U.S. Department of Transportation, Jan. 1985.
22. J. R. Chahine. *Interfacing Databases with Expert Systems: A Retaining Wall Management Application*. Master's thesis. Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh, Pa., Sept. 1986.
23. P. J. Hutchinson. An Expert System for the Selection of Earth Retaining Structures. Master's thesis. Department of Architectural Science, University of Sydney, Australia, 1985.
24. C. T. Hendrickson and B. N. Janson. Expert Systems and Pavement Management. *Proc., Second North American Conference on Managing Pavements*. Toronto, Ontario, Canada, November 2–6, 1987. Ministry of Transportation, Ontario, Canada, and FHWA, U.S. Department of Transportation, 1987, pp. 2.255–2.266.
25. J. J. Hajek, G. J. Chong, R. C. G. Haas, and W. A. Phang. Knowledge-Based Expert System Technology Can Benefit Pavement Maintenance. In *Transportation Research Record 1145*, TRB, National Research Council, Washington, D.C., 1987, pp. 37–47.
26. C. Haas. *Preserver: A Pavement Maintenance Consultant*. Technical report. Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh, Pa., 1986.
27. S. Ritchie, C. Yeh, J. Mahoney, and N. Jackson. A Surface Condition Expert System for Pavement Rehabilitation Planning. *Journal of Transportation Engineering*, ASCE, Vol. 113, No. 2, 1987.
28. R. Duda, J. Gaschnig, and P. Hart. *Model Design in the Prospector Consultant System for Mineral Exploration*. Edinburgh University Press, United Kingdom, 1979, pp. 153–167.
29. R. Banares-Alcantara. *Development of a Consultant for Physical Property Predictions*. Master's thesis. Department of Chemical Engineering, Carnegie-Mellon University, Pittsburgh, Pa., May 1982.
30. M. Rychener. *The DIGR Rule-Based Diagnostic Kernel*. Technical report. Engineering Design Research Center, Carnegie-Mellon University, Pittsburgh, Pa., April 1987.
31. C. Hendrickson and T. Au. *Project Management for Construction*. Prentice-Hall, New York, 1988.
32. C. L. Forgy. *OPS5 User's Manual*. Technical Report CMU-CS-81-135. Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pa., July 1981.
33. L. Browston, R. Farrell, E. Kant, and N. Martin. *Programming Expert Systems in OPS5: An Introduction to Rule-Based Programming*. Addison-Wesley, Inc., Reading, Mass., 1985.
34. D. E. Weatherby. *Tiebacks*. Research Report FHWA/RD-82/047. FHWA, U.S. Department of Transportation, July 1982.
35. R. O. Duda, P. E. Hart, and N. J. Nilsson. Subjective Bayesian Methods for Rule-Based Inference Systems. *Proc., AFIPS 1976 National Computer Conference*, 1976, pp. 1075–1082.

Development of Expert Systems Technology in the California Department of Transportation

STEPHEN G. RITCHIE, LOUIS F. COHN, AND ROSWELL A. HARRIS

This paper presents the results of a research project the principal objective of which was to prepare a plan for the development and implementation of knowledge-based expert systems (KBES) projects throughout the California Department of Transportation (Caltrans). A major part of the project involved development of a special questionnaire and conduct of 50 in-depth interviews of Caltrans senior managers and engineers to identify candidate projects. Forty-five such projects were identified and ranked by priority. In addition, the following aspects were addressed: resource and time requirements for each KBES project, hardware and software needs to best accommodate implementation throughout Caltrans, and recommendations for training. Caltrans has now begun implementation of the plan developed in this research. A major new KBES research project on hazardous waste management has been initiated and is also discussed in this paper.

In June 1987 the California Department of Transportation (Caltrans) signed a research contract with the Institute of Transportation Studies, University of California, Irvine (UCI) to identify and rank by priority potential applications for knowledge-based expert systems (KBES) within Caltrans. The work specified in the contract was performed by civil engineering professors at UCI and the University of Louisville.

The principal objective of the research was to prepare a long-range plan for the development and implementation of candidate Caltrans expert systems projects throughout Caltrans's functional areas. This plan, in addition to identifying and ranking the candidate projects, would identify resource and time requirements for these projects, hardware and software needs to best accommodate departmentwide implementation, and recommendations for training.

Although this research project was constrained in several ways, to the best of the authors' knowledge it is the first of its kind conducted for a state department of transportation in the United States. Therefore for those who might be interested, this paper presents the results of that research effort.

Caltrans has now begun implementation of the plan developed by this research effort and initiated a major new KBES research project with UCI; the current status of a new project on hazardous waste management is also discussed.

INTERVIEWS

A central component of the research involved conduct of a series of in-depth interviews of select senior managers and engineers at Caltrans. The purpose of these interviews was to identify a spectrum of potential KBES applications within Caltrans for evaluation by the researchers. Caltrans research staff selected approximately 50 respondents. Most were division, office, or branch chiefs, and nearly all worked at Caltrans headquarters in Sacramento, where most of the interviews took place during a 1-week period in September 1987. Several district staff were involved in the interviews (Caltrans has 12 districts); in fact, ultimately one of the two highest-ranked candidate KBES projects was initiated by a district.

A specially developed questionnaire (Figures 1 and 2) was administered by each interviewer to guide the discussion. Before the interview each respondent was provided several pages that described the purpose of the research project and the interview and introduced basic KBES concepts. A copy of the questionnaire was also attached. Each information packet contained a cover letter from Caltrans that underscored the importance of the research project and confirmed the respondent's meeting time and place with one of the interviewers.

Most respondents arrived at the meetings with at least one potential application in mind. This enabled Part 1 of the questionnaire to be completed relatively quickly. A separate copy of Part 2 was then completed for each proposed application. In general, no more than two applications could be discussed within the interview hour, although some interviews lasted 2 hr or led to a second meeting.

S. G. Ritchie, Department of Civil Engineering and Institute of Transportation Studies, University of California, Irvine, Calif. 92717. L. F. Cohn and R. A. Harris, Department of Civil Engineering, University of Louisville, Louisville, Ky. 40292.

Name__________ Date __________

Position__________ Start time__________

Division__________ Int. by SR__________

Office__________ LC__________

Branch__________ RH__________

Phone__________ __________

1. (Introductions; complete information above)

   The purpose of this meeting is to help identify high-potential applications of expert systems for the Department. In other words, to identify opportunities to automate and preserve scarce knowledge, and assist Caltrans engineers and staff maximize their effectiveness.

   (Give brief overview of ES, if necessary)

   To begin, could you describe your responsibilities and the function of the unit (Division/Office/Branch) you direct?

2. *OK, let's now consider any issues, both existing and future, that you think the Department faces, and that expert systems technology could contribute to. Do you already have some ideas or proposals for specific tasks expert systems could address?*

   ___ *Yes* (List below)

   ___ *No, problems do exist, but no potential ES applications* (discuss problems to try to identify ES applications below)

   ___ *No, no problems, no ES applications* (probe deeper on problems, or end interview)

Proposed Systems	Problem Priority
1. __________	______
2. __________	______
3. __________	______
4. __________	______
5. __________	______

(probe for more, then seek priority rankings)

Good, now I'd like to discuss each of these in some more detail (go to Part 2).

FIGURE 1 Caltrans expert systems questionnaire, Part 1.

CRITERIA FOR DETERMINING PRIORITIES

Three criteria were used to determine the relative priority of implementation for each candidate KBES project. The criteria were used qualitatively by the researchers and are discussed in the following sections.

Potential Savings to Caltrans

Some expert systems, if properly implemented, could result in significant savings of funds as a result of reduced manpower needs, better use of materials, and so on. Those candidate projects with the highest potential for savings, as judged by the researchers, received the highest priority in this category.

Potential for Straightforward Implementation

Candidate KBES projects that were assessed to be easiest to implement received the highest priority in this category. Among the factors used in this assessment were clarity of the problem definition and solution options, cohesiveness of the knowledge base, and availability of expertise.

Potential for Active Support in Caltrans

For a KBES project to be successfully implemented, it is important to involve decision makers and experts within Caltrans who understand the nature of the development of expert systems and are enthusiastic about the project.

Name____

Proposed System____

1. Could you describe the nature of the problem or task this system would address (an example may be helpful)?

2. Is this a current/increasing/future problem?________________

3. Should this system be PC-based?________________________

4. What would be the system outputs and uses?

5. What would be the user inputs?

6a) Who would be the potential users (e.g., HQ/District/Locals)?

b) How many users?________________________

c) Are computers available to them now?________________________

d) Do they use them; if no, why?________________________

		YES	NO
7a)	Would you say that there is a substantial payoff or benefit to the Department in solving this problem?________________	___	___
b)	Are experts in the Department being lost due to retirements, transfers or promotions?________________	___	___
c)	Are experts scarce because demand exceeds their availability?________________	___	___
d)	Is expertise needed in many locations?	___	___
e)	Is it needed in hostile, expensive or dangerous environments?________________	___	___

8a) What is the extent of algorithmic or mathematical analysis required to get a solution to this problem?

____

		YES	NO
b)	Does the problem require only cognitive, not physical skills?	___	___
c)	Does it require expert knowledge rather than a lot of common sense to get a solution?	___	___
d)	Do experts exist who can solve this problem?	___	___
e)	Do they agree on the solutions?	___	___
f)	Can they articulate and explain their solution methods?	___	___
g)	Can they solve the problem in hours, not days or weeks?	___	___
h)	Do they really understand the problem, or is basic research on the problem required first?	___	___

9a) How many experts exist to solve this problem?

____

b) Where are they?________________________

c) How many are retired experts who could return as consultants?________

FIGURE 2 Caltrans expert systems questionnaire, Part 2.

ASSIGNMENT OF PRIORITIES FOR IMPLEMENTATION

The three criteria defined previously were used to place each candidate KBES project in one of four categories. The four categories are

0: Recommended for immediate implementation
1: Recommended for implementation within 1 yr
2: Recommended for implementation after 1 yr
3: Recommended for future implementation

Only those projects with the highest potential for savings, straightforward implementation, and Caltrans support were placed in Category 0. They are projects that Caltrans should seek to implement immediately as a means to demonstrate the utility of KBES. This category is limited to the two most significant projects identified in the study: Hazardous Waste Site Characterization and Disaster Planning and Management.

Category 1 projects have potential payoffs significant enough to warrant consideration of their development as soon as possible (ideally within 1 yr). Projects in Categories 2 and 3 are worthy of implementation but should be developed by Caltrans personnel who have become proficient in expert systems development through experience they have gained assisting consultants to implement projects in Categories 0 and 1.

PROJECTS RANKED BY PRIORITY

From the interviews and procedures described, 45 candidate KBES projects were identified and assigned to each of the four implementation categories, as follows (listings within categories are not ranked by priority):

Category 0 (2 projects)

- Hazardous Waste Site Characterization
- Disaster Planning and Management

Category 1 (6 projects)

- Pavement Rehabilitation Project Development
- Design Standards Exceptions Advisor
- Hazardous Waste Mitigation Options Advisor
- Incident Traffic Management
- Highway Planting Project Design Advisor
- Assessing Effectiveness of Traffic Mitigation Strategies

Category 2 (22 projects)

- Route Location Study Advisor
- Route Concept Report Advisor
- Sections 16(b) & 18 Advisor
- Regional Transportation Plan Evaluation Advisor
- Financial Data & Trend Interpreter
- Transportation Permit Advisor
- Encroachment Permit Advisor
- Safety Hardware Advisor
- Hazardous Waste Site Evaluation Advisor
- Leaking Underground Fuel Tank Advisor
- Security Analysis Advisor
- Revegetation/Erosion Control PS&E Advisor
- Visually Assessing Highway Projects Advisor
- Bid Pattern Interpreter
- ROW/Utilities Interaction Advisor
- Transit Capital Improvement Project Ranking Advisor
- Accident Analysis Advisor
- Hydrologic Analysis Advisor
- Hydraulic Analysis Advisor
- Water Management Advisor
- Vegetation Control Advisor
- Railroad Relocation Advisor

Category 3 (15 projects)

- Scenic Resource Evaluation Advisor
- STIP/Obligation Plan Development Advisor
- Technology Transfer to Local Agencies
- Equipment Repair Advisor
- Software Selection Advisor
- Traffic Operations Center Advisor
- "Landscaped Freeway" Status Advisor
- Incident Response Advisor
- Traffic Signal Operations Advisors
- Transit Network and Operation Planning Advisor
- Impact Assessment Advisor
- Signal Timing Advisor
- Utility Policy and Procedures Advisor
- Concept Development Advisor
- Environmental Planning Advisor

Summary descriptions of all candidate projects were developed from the interviews and assessments by the researchers. Sample descriptions are presented for projects in Categories 0 and 1 (excluding those related to hazardous waste management, which are discussed later; see "Current Implementation Status").

Disaster Planning and Management (District 7, Los Angeles, Construction and Maintenance)

Problem

The recent earthquake in October 1987 and forecasts of a devastating quake in Southern California, together with the apparent increase in transport of hazardous materials, underscore the need for effective disaster planning for the state highway network. Following a major disaster, the primary responsibility of Caltrans is route recovery. The ability to move personnel and supplies quickly and safely is critical in responding to disaster. An on-line expert system advisor is particularly useful during the emergency event as well for disaster planning. The system should be capable of isolated operation. It will assist in defining the nature and scope of the problem; determining an appropriate response in terms of available personnel, equipment, and materials; determining the ability of the district (District 7 in this case) to provide continued transportation for people and goods; and coordinating and communicating with other public agencies, as well as the public.

Benefits to Caltrans

After a major disaster, Caltrans is a key respondent and manages route recovery and all damage intelligence for state transportation facilities. An expert system can provide a systematic framework for processing information and developing rational responses. It can advise on setting global response priorities for which parts of the transportation system should be reconstituted the fastest, optimize the deployment of forces to particular problem areas, and assist in ensuring that appropriate communication occurs among all involved parties, including district personnel, State Office of Emergency Services, Caltrans headquarters, California Highway Patrol, transit operators, and private contractors.

Comments

This expert system can be used by traffic operations and maintenance personnel, as well as the Executive Office, in District 7 and in headquarters. The system can also serve as a model for other districts.

Pavement Rehabilitation Project Development (Caltrans, Division of Project Development)

Problem

Pavement maintenance and rehabilitation under contract are currently costing the department approximately $200 million per year. Rehabilitation needs and costs will likely increase dramatically in coming years. Many district project design engineers are relatively inexperienced in pavement rehabilitation and lack the expertise to make optimal selections among rehabilitation structural alternatives and develop a complete construction project. An expert system is needed to interpret and assess pavement condition data (both from the pavement management system and from the most recent field inspections), local material availability and cost, material characteristics and properties, and environmental conditions. The expert system can provide assistance and guidance to district project design engineers in engineering the "total solution," including consideration of shoulder treatments, drainage needs, digouts, crack sealing, patching, preleveling, guardrails, superelevation corrections, and ride quality, in addition to the overlay thickness.

Benefits to Caltrans

Implementation of this system will permit scarce, but high-level, Caltrans expertise in pavement rehabilitation to be available to district project design engineers, consultants, and local city and county engineers. There are estimated to be hundreds of potential users within Caltrans alone. More uniform and consistent rehabilitation decision making can be expected to result from use of this system. Caltrans's limited resources to respond to pavement rehabilitation needs can be maximized and the most cost-effective projects can be constructed. Productivity gains in project throughput can also be expected.

Comments

As Caltrans tries increasingly to contract out design work to consultants, many of whom are inevitably unfamiliar with Caltrans practices and procedures, this system assumes greater significance.

Incident Traffic Management (Caltrans, Division of Transportation Operations and Toll Bridges)

Problem

More than 50 percent of freeway delay in California is caused by freeway incidents. Furthermore, this delay is increasing 15 to 20 percent per year. Systematic procedures are required to advise field personnel on incident response and traffic management decision making during freeway incidents. Computerized traffic operations centers are now in use, and more are planned for the state's major urban areas. Although such centers may be able to detect possible incidents, incident verification, incident response, and traffic management decisions are made by using human judgment and experience. There is a need to develop an expert system tool for preincident training of field personnel (i.e., Caltrans traffic operations and maintenance personnel and California Highway Patrol officers), as well as for decision-making advice and backup during actual incidents (possibly using the system on either a desktop or laptop PC).

Benefits to Caltrans

Implementation of this system will improve the ability of Caltrans and associated field personnel to respond optimally to freeway incidents, thus minimizing traffic disruption and improving safety and level of service. More effective use of Caltrans equipment and personnel can also be expected.

Comments

As a training tool, the expert system allows users to develop expertise in at least three simulated phases of incident management: at the traffic operations center, verifying the incident and deciding whether to roll equipment, or, for example, only using ground-mounted signs, rolling to the scene, monitoring new information on the incident and the current and likely traffic impact and potentially modifying response strategies at the site and traffic management strategies upstream and around the site; and finally, at the scene, making operational decisions on response and traffic management strategies, including detour routes and need for signal timing changes, interacting and communicating with other agencies and the public, and coordinating the overall effort. In addition, this expert system will be useful in responding to earthquake damage-induced incidents on the state highway system and can form part of a more extensive disaster planning and response system.

Highway Planting Project Design Advisor (Caltrans, Division of Project Development)

Problem

The workload in highway planting project design is such that consultants are likely to be more involved than ever before. There is a need to provide training to these consultants, as well as to the 40 to 50 Caltrans employees currently involved in this activity. The problem is to develop an expert system that structures the planting design process to make it more uniformly applied by all designers. The tutorial in this expert system will be very valuable. Among other things, the expert system would advise on field checking, plant selection, and irrigation design.

Benefits to Caltrans

Implementation of this expert system will produce uniform and comprehensive highway planting designs. It can expedite and improve the landscape architecture of each highway project and significantly affect maintenance costs through improved design performance and plant selection.

Comments

The financial implications can be very significant, because maintenance costs could be appreciably reduced. In addition, staff time can be saved through more efficient design and review.

Assessing Effectiveness of Traffic Mitigation Strategies (Caltrans, Division of Transportation Planning)

Problem

The objective of this research is to develop an improved decision-making tool to identify and assess transportation systems management (TSM) strategies for mitigating the traffic impacts of new community developments. This tool is an expert system and can be used by state and local agencies, and probably also private developers, to achieve a standardized evaluation framework for mitigation strategies. Local agencies throughout California are increasingly turning to packages of TSM strategies to permit continued development and economic growth to occur while ameliorating level of service and environmental impacts on adjacent state freeway and arterial corridors. However, limited guidance is available to those wishing to perform such analysis because relatively few individuals have the expertise. Also, existing methods and models are limited in scope (particularly their ability to suggest appropriate packages of TSM strategies) and ease of use, which inhibit their implementation.

Benefits to Caltrans

The state of California will benefit from this research through enhanced ability to maintain adequate levels of service in freeway and arterial corridors. This will assist in promoting economic development while also achieving other local and regional social and environmental goals. Expert systems will enable Caltrans to analyze major development proposals and identify appropriate mitigation measures and transportation improvements to alleviate impacts on the transportation system throughout the state.

Comments

As Caltrans attempts to expand its relationship with local government and the private sector in transportation planning and traffic mitigation, tools for consistent and user-friendly quantification of TSM strategy impacts will become essential. The decision-making process to identify appropriate and effective TSM strategies will require growing partnership among the state, local, and private sectors. Caltrans will increasingly be requested to quantify mitigation effectiveness. As economic development occurs and the partnership in planning is formed, the urgent need for expert systems will increase.

RESOURCE COMMITMENTS

The introduction of new KBES technology into the department requires the commitment of services, personnel, and equipment. However, in the opinion of the researchers, the expected return on this investment warrants the commitment. The recommendations to Caltrans are as follows.

Services

Because Caltrans had no in-house expertise and experience in KBES, and because candidate projects in Categories 0 and 1 would be cost effective immediately upon implementation, it was recommended that Caltrans retain consultants to develop and deliver working expert systems for those projects. The Category 0 projects should be initiated as soon as possible, and the Category 1 projects should follow within a year. This will not only allow the rapid deployment of high-payoff expert systems but will also provide the training environment in which Caltrans can develop a group of its own personnel qualified to build expert systems. This group can then be called upon to implement the candidate projects in Categories 2 and 3.

Each candidate project in Categories 0 and 1 should take approximately 1½–2 yr to develop and implement. The majority of this effort will be devoted to developing and validating the requisite knowledge bases. It is likely that the delivered expert systems will be microcomputer based, so they will not require significant investments in new hardware or software. For efficiency, consultant contracts to develop and deliver these expert systems should include the attendant hardware and software.

Personnel

Assuming that this effort to integrate KBES technology into Caltrans is successful, there will be significant expert system development activity under way in house within approximately 3 yr. It is likely that Caltrans will want to establish a centralized knowledge engineering group to most efficiently transfer the technology, assist in developing knowledge bases, and maintain quality control. In all probability, this will be a small group—about five professionals. KBES-related activity will also be spread throughout the department.

Hardware and Software

In recognition of stated user needs, the nature of the vast majority of candidate expert systems identified for Caltrans, and the costs of expert system hardware and software, it is recommended that Caltrans initially specify a microcomputer environment for expert systems development and delivery. Microcomputer-based expert systems development software is becoming increasingly powerful and does not require a large investment, particularly if run-time copies of developed systems can be distributed to Caltrans personnel at low cost, as through a run-time license agreement. It was therefore recommended that Caltrans use (as appropriate) a shell such as the EXSYS expert system software development tool for its IBM-PC-compatible computers.

TRAINING OF CALTRANS PERSONNEL

A major component of the contracts for consultant services to develop and implement the candidate projects in Categories 0 and 1 should be training Caltrans staff, to assist the department in developing the in-house expertise and experience discussed previously. This training should be accomplished primarily in a collegial environment instead of a traditional classroom. However, it may be advantageous for the contractors to conduct periodic workshops for Caltrans personnel with significant involvement. Certainly the first such workshop should occur very early in the effort for the Category 0 projects. Caltrans was advised to begin identifying potential staff members as soon as possible. This was to be done on the assumption that certain of these individuals might eventually become part of the knowledge engineering group discussed previously.

CURRENT IMPLEMENTATION STATUS

The Caltrans Office of Advanced Transportation Technologies has now begun implementation of the plan developed in this research project. In August 1988 a major new KBES project, "Expert Systems Development for Hazardous Waste Management," was initiated with the Institute of Transportation Studies at UCI. The project is being conducted by a multidisciplinary team of civil engineering faculty and graduate research assistants from UCI and hazardous waste management specialists from Caltrans. Policy and technical project advisory committees have been established to also bring together specialists from other state agencies.

This new project integrates the several hazardous waste-related candidate projects identified previously and addresses the hazardous waste management needs of Caltrans in the following two areas:

- Initial site assessment and preliminary site investigation for suspected hazardous waste sites on existing and, particularly, proposed rights-of-way (ROW).
- Initial site assessment and preliminary site investigation of leaking underground storage tanks located at Caltrans maintenance stations.

Early in the Caltrans project development process for proposed ROW, site screening investigations are required when site contamination by hazardous waste is suspected. These investigations and, if necessary, remedial actions must occur before ROW acquisition and construction. Because many existing Caltrans projects, particularly in urban areas, involve sites with suspected contamination by hazardous waste, the potential exists for substantial delays in project delivery. Moreover, there is a need to screen future project corridors for potential hazardous waste sites. The possibility of project delays and associated financial penalties must be minimized, and preferably eliminated, through effective site screening investigations.

Headquarters personnel are too few in number and spread too thin to provide the general assistance required by the districts in investigating these sites and also to provide necessary complex and critical problem solving.

Although it is desirable for the districts to act as independently as possible, there is a pressing need to provide assistance to district personnel for the Initial Site Assessment (ISA) of suspected hazardous waste sites in order to determine the significance of the hazardous waste problem, as well as a need for a more detailed Preliminary Site Investigation (PSI). As part of the process, districts need to be able to prepare Task Orders under the statewide Master Consultant Contract so that consultants can undertake studies to confirm the type and extent of contamination and recommend remedial actions. However, many district personnel do not yet have the knowledge or experience to permit preparation of appropriate and effective Task Orders or, more fundamentally, to conduct the ISA on which they are based, let alone undertake the complex task of evaluating the mitigation alternatives. These problems are becoming critical to Caltrans's ability to timely and cost-effectively deliver projects.

In the case of underground storage tanks, Caltrans has approximately 300 maintenance stations throughout California. Most stations typically have two or three fuel tanks, with additional tanks for pesticides. Many of these tanks are known to be leaking. Caltrans stands responsible for potential soil and groundwater contamination resulting from these leaking tanks. The longer the tanks are left leaking, the greater the environmental hazard and the greater Caltrans's potential liability. The Division of Highway Maintenance needs assistance in developing a statewide program for quickly cleaning up these leaking underground storage tanks. As in Project Development, maintenance personnel have a need to develop Task Orders under the statewide Master Consultant Contract so that consultants can undertake studies to confirm the type and extent of contamination and recommend remedial action.

The objective of the new research project is therefore to develop two flexible, computerized KBES that capture the best expertise and experience available to assist Caltrans's staff in hazardous waste site investigations.

The first expert system, SITE (Site Investigation and Training Expert Advisor), will supplement existing headquarters staff, particularly the Hazardous Waste Management Group, which is limited in size and mission relative to the extent of assistance needed by the districts. SITE will function as an expert system colleague (or computerized assistant) in the development of Task Orders for the PSI. The system will also be used as a training tool to tutor both existing and new staff who have hazardous waste-related responsibilities in districts in what needs to be done for both the ISA and the PSI and why it is important.

In preparing Task Orders for PSI, SITE will automate the ISA. This phase of site investigation is of utmost importance, and requires extensive knowledge and judgment. Some of the information to be considered includes:

- Type of problem, for example, landfill, lagoon, leaking tank (under- or above ground), surface spill, contaminated soil, contaminated groundwater, pipe leak.
- Type of Caltrans project and nature of construction.
- Site information (such as published geological and hydrologic records, well logs, geography, historical land use records), for example, what chemicals were used when and where, active or inactive site, sewer or septic tank on site, wells or city water.
- Probable types of contaminant, for example, petroleum fuels, industrial chemicals, dry cleaning chemicals, solvents, metals, asbestos, pesticides.
- Relevant local, state, and federal laws.

SITE will use this information and its reasoning abilities to generate appropriate investigation strategies to confirm the type and extent of contamination. SITE will then present and document these procedures as a PSI Task Order.

The second expert system, TANK (Tank Advisor and Knowledge System), will assist Division of Highway Maintenance staff. TANK will also function as an expert system colleague or computerized assistant in the development of Task Orders for PSI of leaking underground storage tank sites. Although similar in concept to SITE, in automating the ISA and considering the issues in developing a PSI Task Order, TANK is necessarily somewhat different from the project development system. With TANK, the existence of a leak is usually known from precision tank testing or other means, the general location of the leak is known, and the contamination is on Caltrans property and is Caltrans's responsibility. The focus is therefore on investigation and remedial strategies. In contrast, for new projects, Project Development may have the option to change an alignment or not acquire a parcel of land in order to avoid a hazardous waste site. In addition, maintenance stations have various ongoing functions and structures that one wishes to not disturb during site investigation.

Thus, the objective of using TANK is to employ its expertise and reasoning abilities to generate appropriate investigation strategies to confirm the type and extent of contamination from the tank or tanks in question. These strategies include the type of drilling and sampling procedures required and the number, location, and depth of samples, particularly for soil and groundwater. Initial guidance in building TANK will be obtained from the recently developed LUFT (Leaking Underground Fuel Tank) manual developed by the California Regional Water Quality Control Board.

Hardware for development of SITE and TANK consists of Apple Macintosh II microcomputers. The development software is NEXPERT Object. It is expected that working prototype versions of both SITE and TANK will be developed by mid-1989.

CONCLUSION

Presented in this paper are the results of a research project that prepared a plan for the development and implementation of knowledge-based expert systems projects at Caltrans. Forty-five candidate projects were identified and ranked by priority in four categories. To the best of the authors' knowledge, this project was the first comprehensive study of its type conducted for a state department of transportation. The procedures developed should be useful to those considering similar studies in the future, with the caveat that this project was conducted under tight budget and time constraints. However, it is believed that a useful foundation exists that can be readily modified to suit different needs.

Caltrans has begun implementing the plan developed in this research by integrating several of the candidate projects into a major development effort involving two knowledge-based expert systems for hazardous waste management.

ACKNOWLEDGMENTS

This paper is based on research conducted for Caltrans, with additional support from the Institute of Transportation Studies, University of California, Irvine, and published in a final report (*1*). The assistance of these agencies and the numerous individuals from Caltrans who participated in this research is greatly appreciated.

REFERENCE

1. S. G. Ritchie, L. F. Cohn, and R. A. Harris. Long-Range Plan for Development and Implementation of Artificial Intelligence/Expert Systems Technology at Caltrans. Final report. Institute of Transportation Studies, University of California, Irvine, 1988.

Development of an Expert System To Assist in the Interactive Graphic Transit System Design Process

NATARAJAN JANARTHANAN AND JERRY B. SCHNEIDER

Urban public transit plays a vital role in the functioning of most urban areas. Transit network design is one of the important components of urban transit planning. Because of the complexity of planning urban transit, the components can be best handled by using interactive graphic methods, coupled with ways to evaluate alternative transit network designs from multiple conflicting criteria. The research presented in this paper focuses on an investigation of the applicability of a knowledge-based expert systems (KBES) approach to increasing the productivity of the transit network design process. For this research, an interactive KBES, TNOP_ADVISOR, was developed to assist in the development of high-performance transit network designs. TNOP_ADVISOR provides advice about how to modify designs to obtain improved performance. A network simulation software package, Transit Network Optimization System (TNOP), provides the capability for modifying and predicting the performance of these designs. Three tests are conducted using the interactive KBES and multicriteria evaluation capability as support tools for the conduct of a TNOP-based design process. The results show that the advice provided helped produce high-performance designs in all cases.

Urban public transit usually plays a major role in the efficient functioning of most urban areas. To play this role properly, the transit system should be planned and operated in the most efficient manner possible within the many existing constraints. Transit planning consists primarily of transit network design, network evaluation, and run cutting. Transit network design involves the design of routes, headways, layover times, and departure times, which together determine patronage and loading patterns. Traditionally transit network design work has been done using either heuristics or mathematical optimization techniques (*1–9*). In general, optimization techniques are applicable only to small networks. Heuristic methods are usually required for large networks because of the heavy computer time requirements of optimization techniques.

Heuristic interactive graphics methods that allow online interaction between the user and the machine have been successfully employed recently in transit network design work (TNOP, EMME/2). Heuristic methods employ empirically derived rules for a systematic search for near-optimal solutions. Interactive methods use human intuitive capabilities and knowledge to help search for the best solution. But such methods depend on the knowledge and experience of a very capable user. Such persons are rare. For this reason, interactive graphics may not always produce high-quality designs. The recently developed knowledge-based expert systems (KBES) approach involves capturing the domain knowledge of one or more experts and using it to structure a knowledge base that can then be consulted to obtain good advice about how to solve particular problems. This approach makes the expertise of a few available to many and can substantially improve the productivity of the design process. KBES are widely used in medicine, engineering, and other fields, but they have not yet been used in transit network design. This research is designed to develop and apply the KBES methodology as an aid to the designer who wishes to make more productive use of the Transit Network Optimization System (TNOP) software.

The major task of this research is to translate the knowledge and experience about designing transit systems, gained by the authors over several years of using TNOP for research and instructional activities, into a well-structured knowledge base. When this is accomplished, less experienced and knowledgeable persons can make much more productive use of TNOP's many capabilities. The task is not to fully automate the design process but to provide a significant consultation capability that will make available a large quantity of expertise to anyone who wishes to use it.

TNOP

Overview

TNOP consists of a large set of computer programs that are used to design and simulate the performance of alternative bus and rail transit systems. An overview of TNOP and its components is shown in Figure 1 (*10*). The system is designed to analyze fixed-route, fixed-schedule transit systems. It provides easily understood graphics displays

N. Janarthanan, Puget Sound Council of Governments, 216 First Avenue South, Seattle, Wash. 98104. J. B. Schneider, Department of Civil Engineering, University of Washington, Seattle, Wash. 98195.

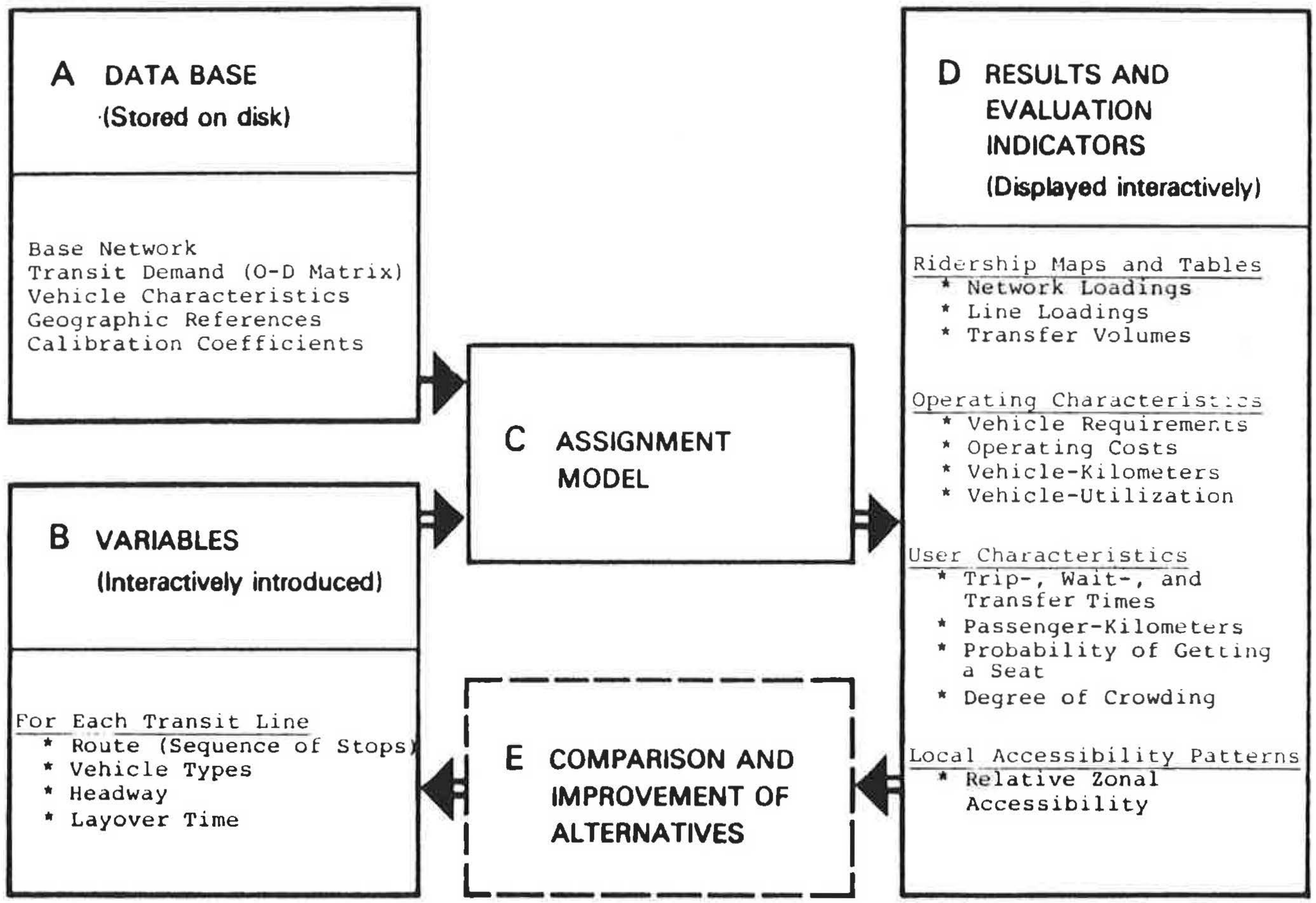

FIGURE 1 Overview of TNOP.

that can be accessed via a user-friendly interface. The interactive graphics system allows the transit planner to examine a wide range of design alternatives and compare their performance to find successively better ways of providing transit services that closely match a particular origin and destination demand pattern.

TNOP was developed in Switzerland by Rapp and others in the late 1970s. The PRIME version is used by the Civil Engineering Department at the University of Washington. The input-output functions and interactive modules were modified to suit CYBER requirements. More modifications were made to eliminate a few bugs and to provide easier access to any menu item in the program. The CYBER version of the program has nearly 100 subroutines and requires 43 working files.

Design Process

The design process supported by TNOP is shown schematically in Figure 2.

Step 1: Input all data and verify them by using the data analysis modules. Use trip desire lines, production and attraction plots, and travel time contour maps to gain familiarity with the data and check for data errors.

Step 2: Define the transit routes, one by one, using the design menus.

Step 3: Specify the service attributes for each route, including headway and the numbers and types of transit vehicles to be used.

Step 4: Assign the trips on the specified transit network.

Step 5: Review the areawide performance values, transit link-load patterns, and other performance measures by using the many tabular and graphics displays. Optimize the headways by comparing the total capacity on the line with the maximum load.

Step 6: If the current design needs changes, repeat Steps 2 through 6.

Step 7: Perform timetable optimization.

By using TNOP, high-performance designs are achieved by a trial-and-error process supported by computer graphics and guided by human intuition as it is shaped by human pattern recognition and other thinking and analytical capabilities. After seeing and thinking about these results, users must select some design modifications that they think will produce better performance measures. This normally involves making some route and service attribute changes. This modified design is then evaluated and the same analysis and modification process is repeated.

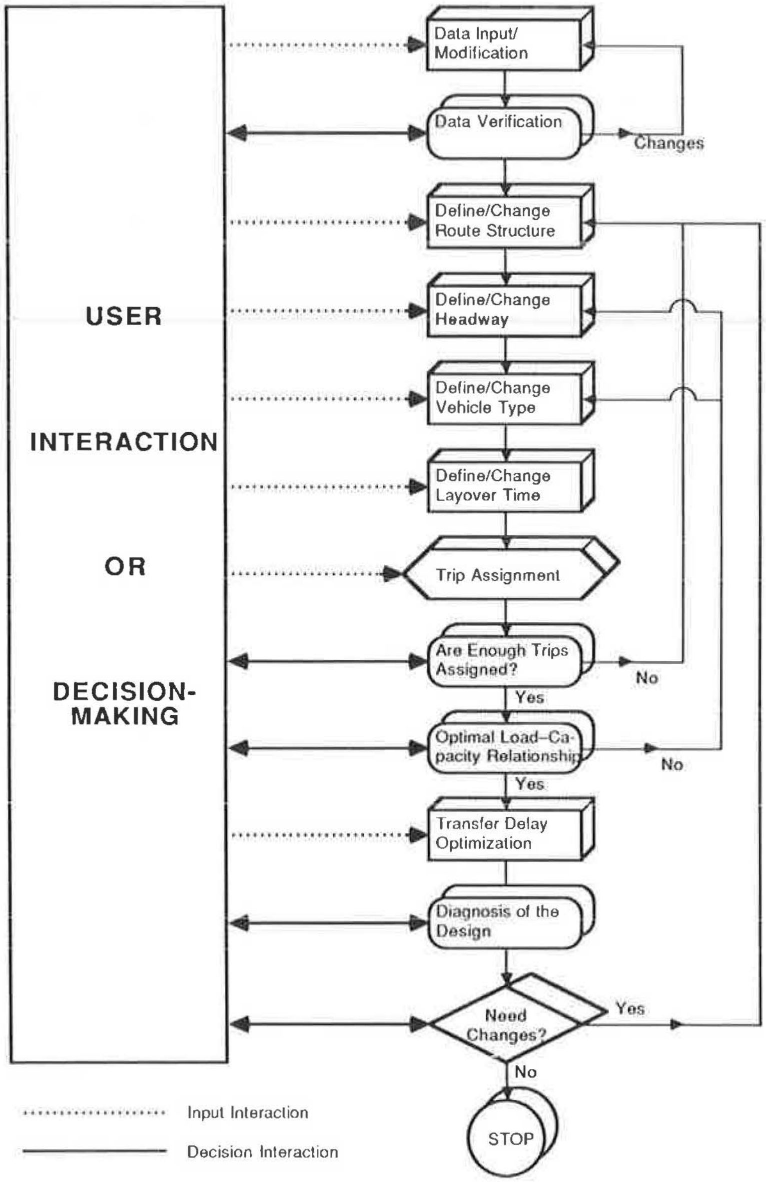

FIGURE 2 Design process supported by TNOP.

Figure 2 clearly indicates the importance of the user's hand in the design process. The user has to interact at every level and make decisions about the input or the next module to be invoked. The user needs to make many decisions, and a user with no or minimal experience will have problems making good decisions. Assisting the user by giving suggestions about some of the input to or analysis of the design will be extremely useful. This will help the user make fast and good decisions that will lead to a design with improved performance. A good KBES can help to achieve this goal.

REVIEW OF KBES

KBES is one of the results of applications of artificial intelligence (AI) research to software programming. Artificial intelligence is a branch of computer science that studies ways of enabling computers to do tasks that appear to require human intelligence (*11*). Expert systems are named for their essential characteristic: they provide advice for problem solving that is derived from the knowledge of experts. Expert systems typically use a set of rules and facts to make inferences that are reported as conclusions. The inference process relies heavily on theories of logical deduction (*12*). The objective of an expert system is to help the user choose among a limited set of options, within a specific context, from information that is more likely to be qualitative than quantitative.

Expert systems have become popular recently and have been successfully applied in many fields. This includes diagnosing infectious diseases [MYCIN (*13*)], finding the structure of chemical compounds [DENDRAL (*14*)], per-forming mathematical symbol manipulations [MACSYMA (*15*)], and exploring for minerals [PROSPECTOR (*16*)]. In addition, expert systems are available to assist in analyzing land use laws and other legal issues (*17*), determine whether a proposed land use meets zoning and other local land use regulations, give advice on building regulations, or estimate probable damage to property in the event of natural catastrophe.

The framework for an expert system primarily consists of the knowledge base, inference engine or inference machine, context or working memory, explanation module, and user interface. Figure 3 shows the components of an expert system. The collection of facts, rules, and computational procedures that represent the domain is called its knowledge base; it is the power base of the expert system. The set of procedures for manipulating the information in the knowledge base to reach conclusions is called the control mechanism, or inference engine. The objective of the inference engine is to find one or more conclusions for a subgoal or for a main goal of the consultation. It searches the facts and rules in the knowledge base and identifies and stores conclusions to use in new facts for subsequent inferencing. The context or working memory contains all the information derived from the inferencing process. This information describes the problem being solved, the rules that have been "fixed," and the conclusions derived from them. The explanation module contains explanations for every inference made or piece of advice given. The user interface provides for a dialogue between human and machine.

Knowledge acquisition and representation are the most difficult parts of building an expert system. They often require the knowledge engineer to interact intensely with one or more experts in the application domain. Not all problems are suitable for expert systems. For successful application, there must be an expert or experts in the domain, and the problem should be specialized. The expert selected must be able to articulate the special knowledge needed to solve problems in the domain. If this problem solving involves the use of rules of thumb or symbolic reasoning, an expert system might be appropriate and helpful.

APPLICABILITY OF EXPERT SYSTEMS TO TRANSIT NETWORK DESIGN

A conventional interactive approach to transit network design using TNOP is shown in Figure 4. A successful design will depend on the knowledge and intuitive skill of the user, who must conduct a cumbersome iterative search process that may or may not produce a better design. Also, this traditional approach is user dependent and lacks consistency and reliability. These problems can be solved by the expert systems approach, by which available design expertise is transformed into a visible format so it can be used and maintained by nonexpert designers.

Transit network design problems are suited to a KBES approach because it is heuristic, has no explicit solution steps, and requires domain knowledge to solve the problem. It takes quite a few years of experience for an average user to build domain knowledge. Developing a KBES to assist in transit network design will help the nonexpert users find high-performance designs quickly.

The expert systems approach captures the knowledge of one or more experts and uses it to solve similar problems to eliminate the user-dependency factor in solving the design problem. Figure 5 shows the role of the expert system in solving transit network design problems. This interactive method uses a knowledge base to identify the ways to improve the network design. The various performance measures and attributes of a transit network design will be passed through the KBES to identify the flaws in the design and generate advice about how to improve it. This iterative process will stop when the user finds a satisfactory transit network design or the KBES can provide no further advice that is feasible within the constraints of the problem.

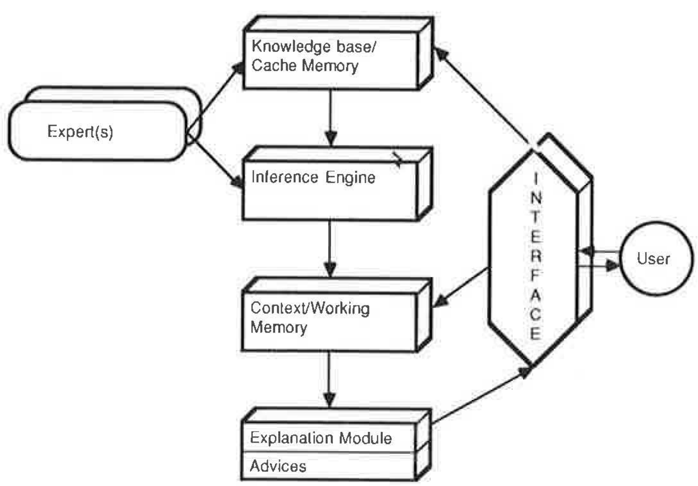

FIGURE 3 Components of an expert system.

INTERACTIVE APPROACH TO TRANSIT NETWORK DESIGN USING TNOP AND TNOP_ADVISOR

This section presents the knowledge-based interactive approach to transit network design. The methodology uses TNOP (a simulation model), TNOP_ADVISOR (a KBES), and CONCORD_NL (a nonlinear concordance analysis method for comparing and ranking alternatives). A software package called PRISM was also developed to select and arrange the results from a TNOP run to enter into TNOP_ADVISOR. The knowledge base consists of many knowledge subbases, which are discussed in detail. The proposed methodology is tested in three different test networks, and the results are reviewed and analyzed.

Development of the Knowledge Base TNOP_ADVISOR

A knowledge base is the main component of an expert system. In this research, the knowledge base includes the knowledge to design an efficient transit network using TNOP. The knowledge base must be structured to use the full design capabilities of TNOP and recognize its constraints. The following section discusses in detail the analysis of design strategy using TNOP.

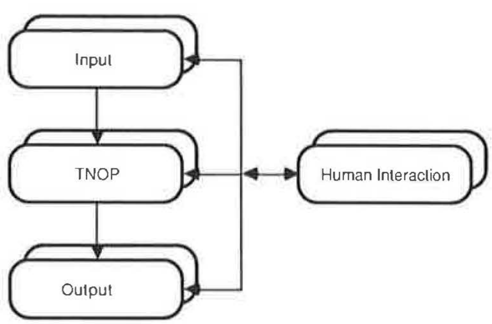

FIGURE 4 Conventional approach to transit network design.

Analysis of Design Strategy Using TNOP

The main inputs for TNOP are the base network, transit mode demand pattern, TNOP model parameters, line definition, and service attributes. For a given design problem the base network, demand pattern, and model parameters are fixed. These fixed parameters are difficult to change without making planning changes. Planning changes normally include changes in routes and are often not possible except for minor extensions or reductions in existing routes. In real-world problems, it is easier to extend or add new routes than it is to shorten or remove routes. The variable parameters in TNOP are mainly service attributes such as headway, vehicle type, layover, and departure times on routes. These variables determine the operational attributes of the transit system. The knowledge base developed in this research focuses on generating improved designs by making only operational changes. The implications of changing these variables on the performance of a transit system will be explained.

Changes in the headway on a route affect the frequency, waiting time, number of vehicles required, operating cost, and route use. One of the constraints attached to modifying headways is that they must satisfy policy headways. The objective of making headway changes is to achieve optimal capacity conditions on a given route in order to match the maximum link load on the route. If a given route has a prespecified policy headway, an optimal headway can be suggested only when it is shorter (i.e., better) than the policy headway. A change in vehicle type can also help to increase the vehicle use on the route. This will have a direct impact on route and system use and operating costs. Changes in layover time will change vehicle requirements and operating costs. Departure time changes will affect the transfer delay performance of the system.

When operational changes are allowed, advice will be given about how to modify one or more of the variable parameters. When planning changes are allowed, general advice will be given about how to modify one or more routes. The knowledge base will be able to handle two situations: when planning changes are allowed and when they are not. A constraint inherent in using TNOP is that a mix of different vehicle types cannot be used on a given route. TNOP provides various statistics and graphic dis-

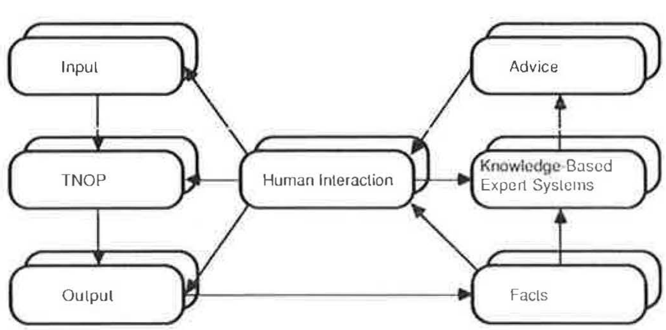

FIGURE 5 Expert system approach to transit network design.

plays to aid in the evaluation of a network design. All the statistics will be analyzed to get an accurate and detailed description of a design for use by the knowledge base, which performs the inferencing task and generates design modification advice.

Description of Expert System Shell RUNNER

To build an expert system, a shell must first be selected. A shell is a general framework for building an expert system. It consists of an inference engine and the necessary utility software to create context and explanation modules when the shell is invoked. The components of a typical shell are shown in Figure 6. For this research, instead of building a new shell or using some commercially available shell, the authors used a shell called RUNNER, developed by Yeh (*18*). The main advantage of using an in-house shell is its flexibility to customize functions. A few functions were added or modified for this study. It also offered a rule-based system for the knowledge base that is easy to use and modify for transit network design applications. RUNNER was written in LISP language.

Structure of TNOP_ADVISOR

The knowledge base that was developed to assist in designing a transit system using TNOP is called TNOP_ADVISOR.

Figure 7 shows the structure of TNOP_ADVISOR. The main control knowledge base is the master controller, which reads the current facts to invoke a correct knowledge base that is properly sequenced. The main knowledge base can invoke any of the 11 knowledge-base submodules.

TNOP_ADVISOR consists of both rule-based and function-based knowledge modules. The main purpose of the functions is to carry out an operation repetitively. A total of 199 rules are included in the knowledge base that are used to evaluate the facts and make inferences. These rules have if-then and conditional formats. A total of 111 items of advice are available to choose from depending on the inferences made by the knowledge base.

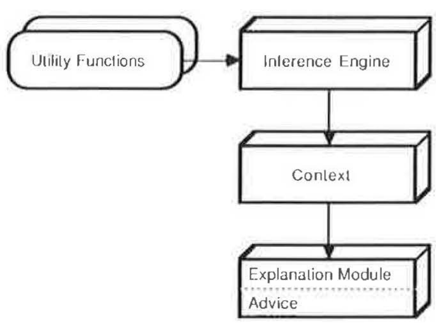

FIGURE 6 Components of a typical shell.

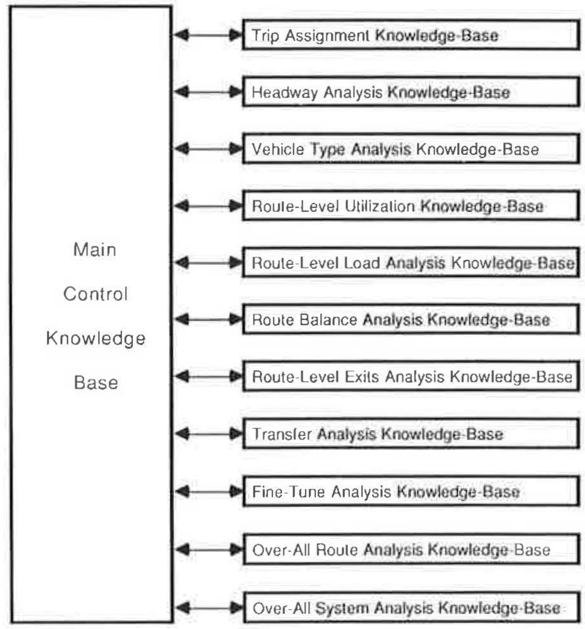

FIGURE 7 Structure of TNOP_ADVISOR.

Of the 11 knowledge-base submodules, only 2 modules (headway and vehicle type analysis) give advice related to operational changes. The other nine modules give advice on planning changes or on combinations of planning and operational changes. The following sections briefly describe each module and its functions.

Main Control Knowledge Base

This is the master controller; it directs the flow and sequence of knowledge-base processing. Its main functions include invoking the correct knowledge base submodules based on the available facts. This submodule also verifies whether planning changes are possible before invoking a module.

Trip Assignment Knowledge Base

This is the first knowledge base submodule that main control will invoke whether planning changes are possible or not. This submodule checks the facts to verify that the number of trips assigned by the current design satisfies the "minimum number of trips to be assigned" condition. All of these conditions or combinations of conditions are compared with the facts and, based on the result, inferences are made and advice is derived.

If the percentage of trips assigned is less than the minimum number of trips to be assigned, then the trip assignment submodule will fail. Depending on the size of the difference and the location of large unserved nodes, the program will identify some unserved nodes that could be connected to the network design to meet the minimum service requirement. If the analysis infers that it has failed, further analysis will stop. Advice will be given about how to redesign the route structure by adding more nodes.

Headway Analysis Knowledge Base

This knowledge base checks the existing headway, vehicle type, and maximum loading on each route. If the maximum loading does not match the capacity provided, the knowledge base will generate advice on what optimum headway to use. It will also check whether the route has a policy headway, and appropriate advice will be given based on whether the optimum headway is less or more than the policy headway. In addition, a layover time will be suggested. If the optimum headway necessary for the given vehicle type is less than 1 min, advice will be given to use a different vehicle type having a much higher capacity. If that is not possible, advice will be given to make planning changes, such as changing the route structure.

If a system has even one route with nonoptimal headway (when there is possible and feasible change on that route), the headway analysis will be considered failed and further analysis will stop. Because further analysis and advice will not be based on optimal headway condition, this approach is followed.

Vehicle Type Analysis Knowledge Base

This knowledge base scans various available vehicle types and their capacities. It also checks each route for the vehicle type used and how closely this matches maximum loading conditions. If another vehicle type can provide optimal capacity conditions, then it will be recommended. Along with the recommended optimal vehicle type, the optimal headway for that vehicle type and layover time will be suggested. The rules reflect the view that providing 90 percent of capacity (slightly underloaded) is better than providing 110 percent (slightly overloaded) in order to optimize resources and allow for better use.

Route Balance Analysis Knowledge Base

This analysis is carried out only for two-way routes. It compares the loadings in one direction along the route with those in the other direction. Consider an example of a two-way route with five links. The link loads in Direction A are compared with the loads in Direction B. The advice is based on the ratio of total link load in one direction to total link load in the other. The ratio is calculated by dividing the smaller load by the larger load. A ratio of 0.9 or higher indicates a very well balanced condition. When the ratio lies between 0.75 and 0.9, the route is judged to have an acceptable balance condition. Previous experience has shown that if the ratio is less than 0.75, the route is poorly balanced and will have very low use and will decrease the overall loading performance of the system.

The purpose of this analysis is to identify two-way routes with directional loading problems. Many times, a two-way route may be efficient in only one direction. This will reduce drastically the average and also systemwide use of the route. The advice ranges from adding more origins and destinations to the weaker direction to converting it into a one-way route.

Average Utilization Analysis Knowledge Base

This analysis is directed toward the identification of routes with below-average use. Above-average use on a route indicates that the route has better loading (link loads) and headways than does the areawide average. An optimal headway gives minimum operating costs in many instances. By identifying the routes with below-average use, advice can be derived about improving it. The advice includes planning changes like adding or deleting links or dropping the route altogether. This knowledge base will not analyze a route if it has a policy headway.

Maximum Load to Average Load Analysis Knowledge Base

Two useful indicators of route performance are maximum load and average load. The average load gives the overall estimate of load conditions on a route, whereas maximum load identifies the maximum load on a particular link in the route being examined. The maximum load governs the optimal headway calculations. When the difference between the maximum load and average load is high, it indicates that the route needs higher capacity to serve only the maximum link, because many other links do not need more capacity. A high ratio indicates that the route is not efficient and needs improvement.

This knowledge base examines the average and maximum loads of each route, and the ratio of maximum load to average load is calculated. If this ratio is greater than 2.25, the route is bad and needs planning changes. It should be redesigned or dropped from the system. If the ratio is less than 1.50, the route is good. If the ratio lies between 1.50 and 2.25, the route is acceptable, but only a few routes can be in this condition. These boundary numbers are derived from previous experience that was used to define the rules in the knowledge base.

Exit Analysis Knowledge Base

The link loads and exit patterns of a route indicate how frequently a route is used. For example, a route with few exits and a large number of heavy link loads is one over which people travel longer distances, compared with routes where there are many exits and few heavily loaded links. From previous design experience with TNOP, it was found that the ratio of total link loads (sum of all link loads) on a route to total number of exits yields a useful indicator of whether the route serves short or long trips.

Transfer Analysis Knowledge Base

A transfer analysis involves examination of the number and spatial pattern of transfers in the system. The number of transfers in a transit system is a valuable performance indicator. A large number of transfers (greater than 1.0 per passenger) indicates that many people have to use more than one route to reach their destination. A small value indicates that routes are laid out well and that many passengers do not have to transfer to reach their destination. But in many cases, to obtain operational efficiency when there is not enough demand, making passengers transfer from one route to another is necessary. This knowledge base looks at the total number of transfers in the system and the permitted (designer specified) number of transfers in the system. Advice is given based on the ratio of actual to permitted number of transfers in the system. If the number of transfers exceeds the required limit, the advice identifies which nodes have the most transfers. Recommendations are also given that such nodes be analyzed in detail to see how transfers could be reduced. This module does not provide specific advice because knowledge about how to achieve transfer reductions is still limited. The user needs to consider many things, because it is necessary to modify the route structures to obtain transfer volume reductions. Making planning changes may be the only way possible to obtain a reduction in transfer volumes at particular locations.

Fine-Tune Knowledge Base

This module provides system-level advice. The analysis includes both average walk time and the ratio of in-vehicle to out-of-vehicle travel time. A high areawide average walk time means that people spend considerable time walking to and from transit stops. The advice varies, depending on whether the average walk time is less than 5 min, between 5 and 11 min, or greater than 11 min.

To study the relationship between in-vehicle and out-of-vehicle time, the ratio of in-vehicle to total travel time (in-vehicle + out-of-vehicle time) is used. The ratio has a theoretical maximum of 1.0. The closer the value is to 1.0, the better the route. From previous design experience, several rules were derived and incorporated into the knowledge base. The advice includes operational changes, like lower headways on routes, or planning changes, like modifications to route locations.

Overall Advice Expert Knowledge Base

There are two knowledge bases in this category. One gives advice on overall route conditions, and the other gives advice on overall system conditions. These modules combine the inferences made by other knowledge bases and make inferences on the overall route and system conditions. Overall route-level advice is based on the inferences from the utilization, maximum load to average load balance, and exits analyses. A route that has good performance for all these indicators gets a good mark on overall condition. Overall system-level advice is given from inferences derived from average use, overall route conditions, and inference results from the fine-tune analysis.

Full-Design Improvement

A user can specify either full design improvement or conditional design improvement as a goal for the knowledge base. In full-design improvement mode, the facts of a given design will be passed through all the knowledge base modules (subject to the conditions of each knowledge base), inferences will be made, and advice will be generated. Figure 8 shows the flowchart of full-design improvement. In full-design improvement mode, the user has the flexibility of skipping any particular knowledge base by specifying appropriate facts and conditions as input before starting the TNOP_ADVISOR. If the focus is on a particular design improvement, the main control knowledge base will execute the knowledge base chosen by the user and then conclude the analysis.

Development of the PRISM Preprocessor

PRISM is a software package that extracts all the facts about a transit design from TNOP output. PRISM needs TNOP output, a demand matrix, and answers to a series of questions from the user. PRISM goes through the TNOP output, extracts the required facts about the design, and

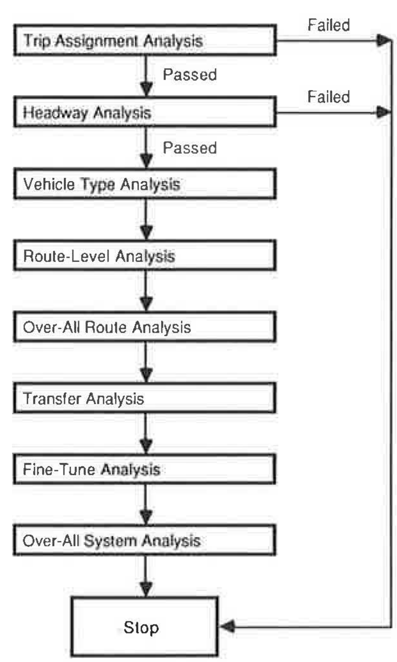

FIGURE 8 Flowchart of full-design improvement.

arranges them properly for reading by the knowledge base. The user needs to answer a few questions interactively about whether the goal is full-design improvement or conditional improvement, whether planning changes are possible or not, and so on. Figure 9 presents the structure of PRISM.

PRISM also analyzes the origin-destination matrix and identifies the origin and destination nodes not currently connected to the network.

CONCORD_NL (Multicriteria Evaluation Software)

CONCORD_NL is a computer program that was written to allow use of the concordance analysis procedure for evaluating alternative transit designs. Transit planning is one of many multicriteria problems that have conflicting goals. This means that better performance for one criterion often cannot be achieved without negatively affecting other criteria values. In addition to these inherent conflicts, the differing opinions of local government agencies, political groups, citizen groups, and system users have to be taken into account. In this research the recently developed CONCORD_NL (*19*) software, which is based on multicriteria evaluation methodology, is used.

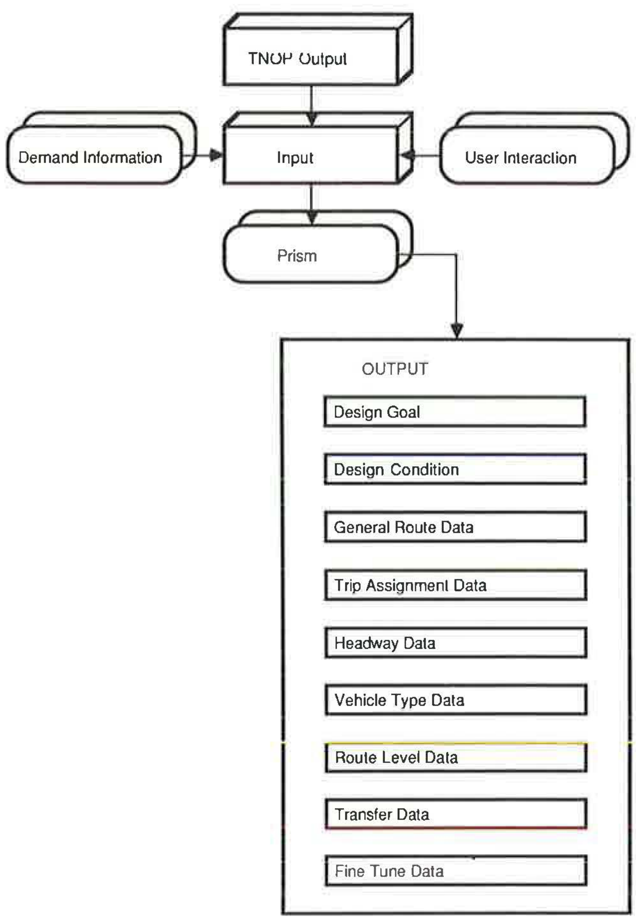

FIGURE 9 Role and output of PRISM.

Concordance analysis is a multicriteria evaluation technique by which alternative plans are evaluated by a series of pairwise comparisons across a set of criteria. It is based on the Electre method originally developed in France. References to and discussions of the development of the Electre and concordance methods are presented by Nijkamp and Van Delft (*20*) and by Giuliano et al. (*21*). CONCORD_NL includes improvements in the normalization procedure by adding a nonlinear normalization method. Figure 10 presents a general framework for evaluating alternative transit network designs.

Testing of TNOP_ADVISOR

The knowledge base TNOP_ADVISOR was tested to evaluate its capabilities to produce advice that, when followed, would produce improved transit network designs. TNOP_ADVISOR was tested on three different design problems. The first two design problems use the same network and demand pattern, and the third problem uses a different network and demand pattern. The first and third problems require that a new transit network be designed. The second problem involves application of TNOP_ADVISOR to an existing network design. The objective of this testing is to prove that TNOP_ADVISOR can handle different problems at different stages of design using different networks.

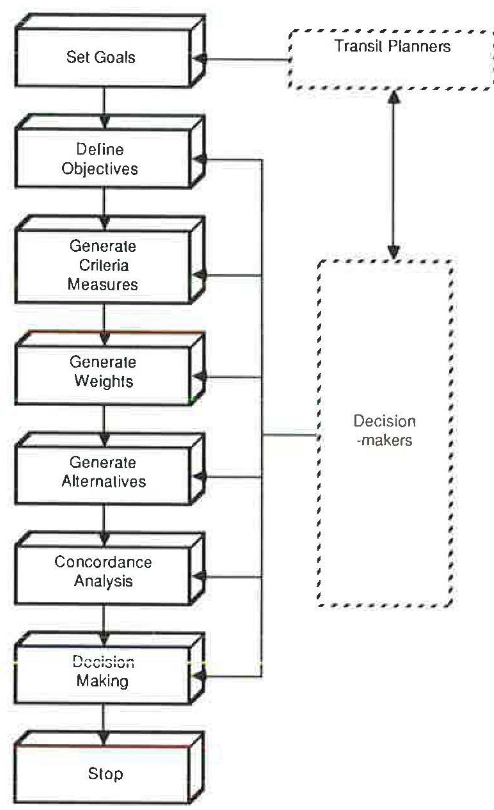

FIGURE 10 Framework for evaluation of transit system improvement alternatives.

In practice, it is often necessary to fine-tune an existing design to improve its performance or to accommodate base network or demand pattern changes, or both, from changes in population or employment.

Figure 11 presents the testing procedure used. The procedure consists of simulating the performance of an initial transit network design, using TNOP as the first step. TNOP can be used either to create a new design or to simulate an existing design. The design process requires the user to input network and demand data, define routes and service attributes, assign trips, and perform timetable optimization. At this point, the TNOP results are fed through PRISM to prepare the facts (performance data) about the design for input into the knowledge base of TNOP_ADVISOR. TNOP_ADVISOR analyzes the design and gives advice about how to improve it. The user can either accept or reject this advice and modify the transit network design using the interactive graphics capabilities of TNOP. The whole process can be repeated until the user is satisfied. Because the transit network design process does not use a mathematical optimization technique, the knowledge base will not decide when to stop; the user decides. The decision to stop is usually made when most of the routes attain satisfactory performance measures and when the advice provided does not refer to operational changes. The knowledge base will give specific advice on route- and system-level operational changes. Advice on planning changes is restricted to indicating the direction for possible changes, and it is up to the user to translate this advice into specific network changes. To show that by following the advice of TNOP_ADVISOR better designs will result, the designs are processed through the CONCORD_NL concordance analysis program, and ranks are computed from the weighted multiple criteria.

Test Problem 1

This problem represents a case in which a net design is to be developed. Figure 12 shows the transit network with node names. This network was designed by Rapp (*10*) as a tutorial network for TNOP software. It consists of 75 nodes and 388 one-way links. Figure 13 displays productions, attractions, and the demand pattern for this problem. The initial design created on this network consisted of five lines (routes). The objectives for this design problem are to serve at least 90 percent of the demand with a system on which the maximum capacity on any link is within 10 percent of the capacity of that route. Full-design improvement is invoked as a goal in the knowledge base, and it is assumed that planning changes are possible. These objectives are used for all the test problems.

The output from the first design (100) from TNOP was passed through PRISM to prepare the facts about the design; then TNOP_ADVISOR was invoked. It analyzed the design facts and used the rules to generate its advice. Design 100 was the initial (first-cut) design. From this starting point, TNOP_ADVISOR's advice was followed exactly to create the next three designs. Concordance analysis results show that each successive design was better than the previous design; this indicates that the advice provided was appropriate and productive. Tables 1 and 2 show the raw performance values, the average dominance ranking for eight weight schemes, and the final rankings for the four designs from the concordance analysis.

Test Problem 2

This design problem uses the same network and demand matrix used previously. In Test Problem 1, a new design

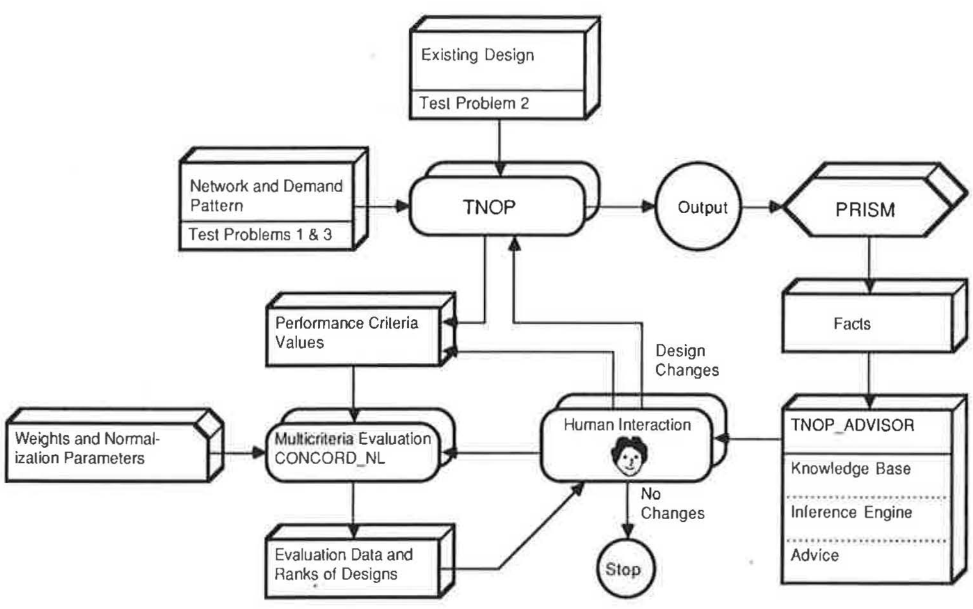

FIGURE 11 Flowchart of procedure for testing TNOP_ADVISOR.

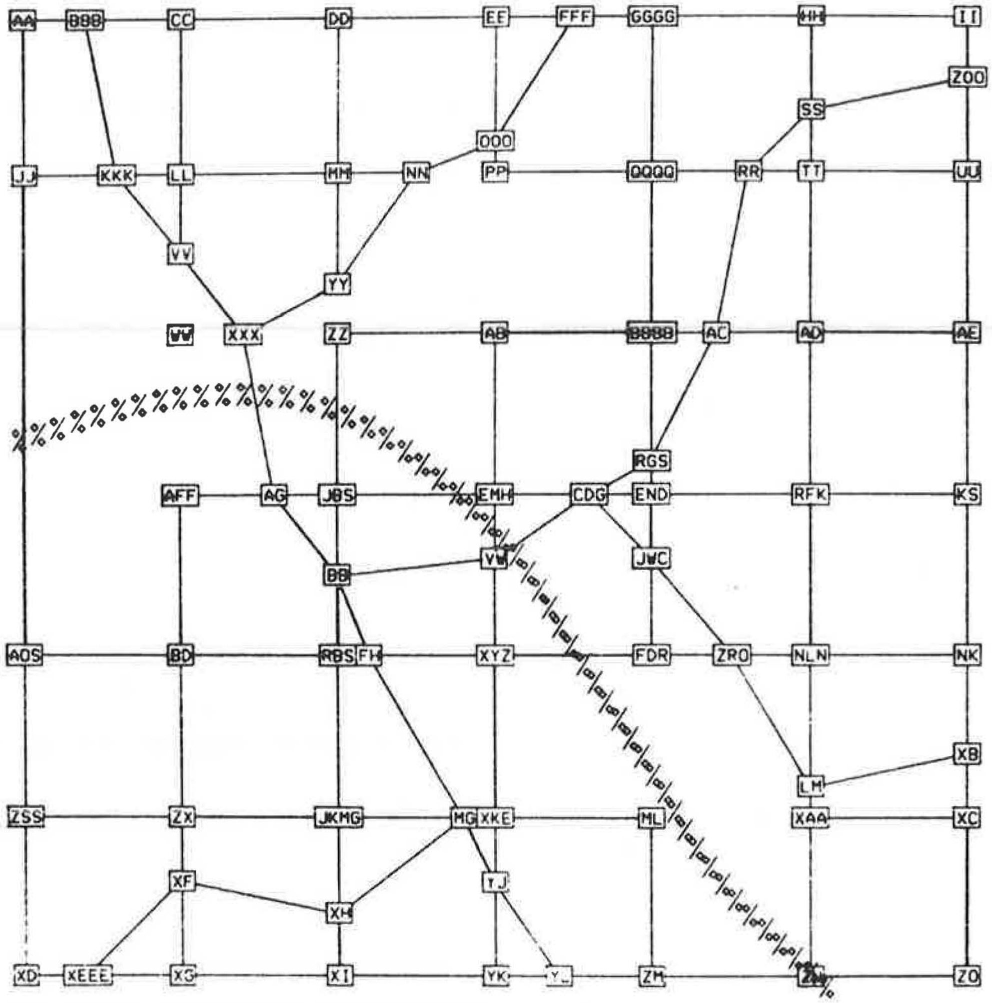

FIGURE 12 Test Problem 1: Base network.

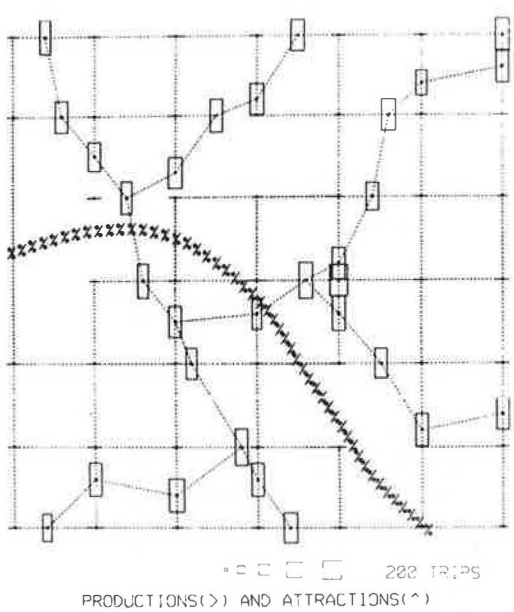

FIGURE 13 Test Problem 1: Demand pattern.

was created. In this problem, an existing design was examined to see if it could be improved. The network design was created by J. B. Schneider, who has had several years of experience with TNOP. It is the best known design developed for classroom use and students are challenged to surpass it. Schneider's design was recreated using TNOP and run through the knowledge base. From this starting point, the advice was strictly followed and the design modified accordingly. It took a total of 14 iterations to reach the final design. Many of the iterations were required to modify headways and vehicle types. Also, the dynamic effect of load changes on the lines caused by headway changes produced many additional iterations.

Concordance analysis was carried out by comparing Designs 100 (Schneider's design), 300, 900, and the final design, 940. Tables 3 and 4 present the raw performance measures for these four designs and the average dominance and final rankings. Design 940 is not a totally dominant one but is better than all others. By following the advice provided, these designs were created and became successively better. If one looks at the performance measures of

TABLE 1 TEST PROBLEM 1: RAW PROJECT EFFECTS MATRIX, DESIGNS 100, 200, 400, AND 500

CONCORDANCE ANALYSIS (A MULTICRITERIA EVALUATION METHOD)

RAW PROJECT EFFECTS MATRIX

	PM		ALTERNATIVES 1 (100)	2 (200)	3 (400)	4 (500)
	1	OPERATING COSTS PER	.120	.130	.300	.290
	2	CAPTIAL COSTS PER PA	1430.000	1588.000	3567.000	3567.000
	3	PERCENT ROUTES WITHI	100.000	100.000	16.600	16.600
*	4	PASS. SERVED / PASSE	2.770	1.760	1.160	1.260
*	5	PASSENGER KILOMETERS	60900.000	68300.000	68600.000	68600.000
*	6	AVERAGE UTILIZATION	142.100	121.400	52.100	53.600
*	7	PERCENT TRIPS ASSIGN	80.600	96.300	96.300	96.300
*	8	TOTAL PASSENGER SPAC	41372.000	53704.000	125863.000	122341.000
*	9	TOTAL ROUTE LENGTH	68.000	87.000	87.000	87.000
*	10	AVERAGE RIDING TIME	.750	.740	.830	.830
	11	NUMBER OF TRANSFERS	1.099	.930	.940	.940
	12	AVERAGE TRANSFER DEL	3.290	3.300	1.440	1.440

PM = PERFORMANCE MEASURES

* = MORE IS BETTER,OTHERWISE LESS IS BETTER

( ) = DESIGN NO.

TABLE 2 TEST PROBLEM 1: RANKING OF ALTERNATIVES, DESIGNS 100, 200, 400, AND 500

AVERAGE DOMINANCE RANKING((CONCORDANCE + DISCORDANCE)/2)

WEIGHTING SCHEMES	ALTERNATIVES 1	2	3	4
1	4.00(0)	3.00(0)	2.00(1)	1.00(1)
2	4.00(0)	3.00(0)	2.00(1)	1.00(1)
3	4.00(0)	2.00(1)	2.50(0)	1.50(1)
4	3.50(0)	1.00(1)	3.50(0)	2.00(1)
5	4.00(0)	3.00(0)	2.00(1)	1.00(1)
6	4.00(0)	3.00(0)	2.00(1)	1.00(1)
7	4.00(0)	3.00(1)	1.50(1)	1.50(1)
8	3.50(0)	1.50(1)	3.50(0)	1.50(1)
TOTAL	31.00	19.50	19.00	10.50

(1)--NON-DOMINATED; (0)--DOMINATED

FINAL RANKING OF ALTERNATIVES

RANK	ALT	DESIGN NO.
1	4	500
2	3(*)	400
3	2(*)	200
4	1(*)	100

(*)--THIS IS NOT A TOTALLY NON-DOMINATED ALTERNATIVE

TABLE 3 TEST PROBLEM 2: RAW PROJECT EFFECTS MATRIX, DESIGNS 100, 300, 900, AND 940

CONCORDANCE ANALYSIS (A MULTICRITERIA EVALUATION METHOD)

RAW PROJECT EFFECTS MATRIX

	PM		ALTERNATIVES 1 (100)	2 (300)	3 (900)	4 (940)
	1	OPERATING COSTS PER	.320	.310	.250	.240
	2	CAPTIAL COSTS PER PA	2953.000	2821.000	2646.000	2597.000
	3	PERCENT ROUTES WITHI	50.000	12.500	.000	.000
*	4	PASS. SERVED / PASSE	1.900	1.980	2.230	2.270
*	5	PASSENGER KILOMETERS	45900.000	45900.000	46200.000	46200.000
*	6	AVERAGE UTILIZATION	42.200	43.900	48.200	49.300
*	7	PERCENT TRIPS ASSIGN	95.600	95.600	95.600	95.600
*	8	TOTAL PASSENGER SPAC	105316.000	101086.000	92711.000	90727.000
*	9	TOTAL ROUTE LENGTH	134.000	134.000	134.000	134.000
*	10	AVERAGE RIDING TIME	.799	.795	.770	.770
	11	NUMBER OF TRANSFERS	.559	.557	.516	.514
	12	AVERAGE TRANSFER DEL	1.471	1.604	2.280	2.275

PM = PERFORMANCE MEASURES

* = MORE IS BETTER,OTHERWISE LESS IS BETTER

( ) = DESIGN NO.

TABLE 4 TEST PROBLEM 2: RANKING OF ALTERNATIVES, DESIGNS 100, 300, 900, AND 940

AVERAGE DOMINANCE RANKING((CONCORDANCE + DISCORDANCE)/2)

WEIGHTING SCHEMES	ALTERNATIVES 1	2	3	4
1	1.00(1)	2.50(1)	4.00(0)	2.50(0)
2	2.00(1)	2.00(1)	4.00(0)	2.00(0)
3	4.00(0)	3.00(0)	2.00(1)	1.00(1)
4	4.00(0)	3.00(0)	2.00(1)	1.00(1)
5	1.00(1)	3.50(0)	3.00(0)	2.50(0)
6	4.00(0)	3.00(0)	2.00(0)	1.00(1)
7	4.00(0)	3.00(0)	1.50(1)	1.50(1)
8	4.00(0)	3.00(0)	2.00(1)	1.00(1)
TOTAL	24.00	23.00	20.50	12.50

(1)--NON-DOMINATED; (0)--DOMINATED

FINAL RANKING OF ALTERNATIVES

RANK	ALT	DESIGN NO.
1	4(*)	940
2	3(*)	900
3	2(*)	300
4	1(*)	100

(*)--THIS IS NOT A TOTALLY NON-DOMINATED ALTERNATIVE

Designs 900 and 940, Design 940 has optimum vehicle types and headways on all lines, whereas Design 900 does not have optimal vehicle types on lines 1 and 2. Design 940 is cheaper and has better performance measures than Design 900. The results from this test problem also prove that TNOP_ADVISOR can be useful to fine-tune and improve an existing design as well as create a new design.

Test Problem 3

The network used for this problem is different from the other two. Figure 14 shows the network and node names. This hypothetical city consists of a hub-and-spoke street system. The network has 97 nodes and 176 two-way links. Compared with the previous problem, this one is much bigger, with a total of 41,730 trips in its transit demand matrix. Figure 15 presents the demand pattern in this network. One of the main objectives in using this problem is to search for evidence that TNOP_ADVISOR can be useful and give appropriate advice regardless of the network, demand pattern, or route pattern used.

It took only four iterations to find a high-performance solution to this problem. Concordance analysis was conducted comparing Designs 100, 200, 300, and 400. Design 200 is better than Design 100, and Design 300 is better than Designs 100 and 200. Design 400 is a nondominant design that is better than the three previous designs. Table 5 presents the performance measures of Designs 100, 200, and 400. Table 6 presents the average dominance and final ranking of the four designs.

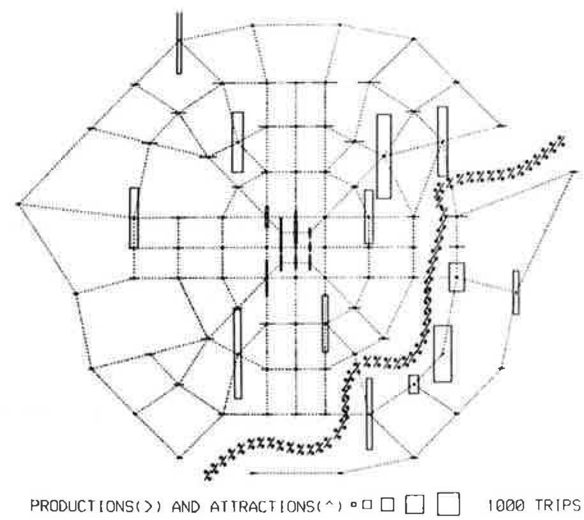

FIGURE 15 Test Problem 3: Demand pattern.

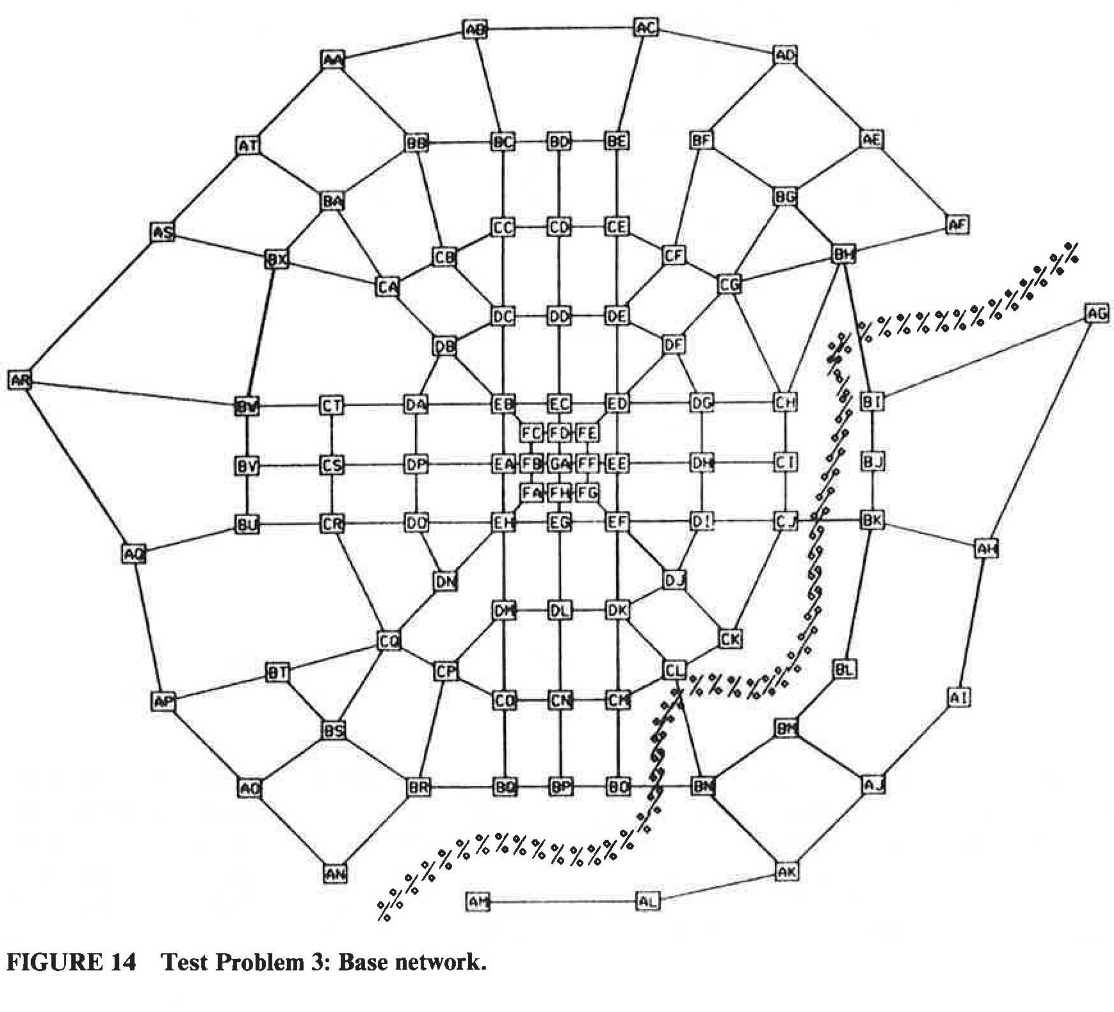

FIGURE 14 Test Problem 3: Base network.

TABLE 5 TEST PROBLEM 3: RAW PROJECT EFFECTS MATRIX, DESIGNS 100, 200, 300, AND 400

CONCORDANCE ANALYSIS (A MULTICRITERIA EVALUATION METHOD)

RAW PROJECT EFFECTS MATRIX

PM		ALTERNATIVES 1 (100)	2 (200)	3 (300)	4 (400)
1	OPERATING COSTS PER	.250	.270	1.000	.950
2	CAPTIAL COSTS PER PA	563.000	620.000	2128.000	2090.000
3	PERCENT ROUTES WITHI	100.000	100.000	18.000	27.000
* 4	PASS. SERVED / PASSE	61.600	5.430	1.460	1.510
* 5	PASSENGER KILOMETERS	649200.000	705100.000	695300.000	695500.000
* 6	AVERAGE UTILIZATION	186.400	183.700	47.800	49.600
* 7	PERCENT TRIPS ASSIGN	76.600	92.700	92.700	92.700
* 8	TOTAL PASSENGER SPAC	289007.000	377316.000	1430051.000	1378550.000
* 9	TOTAL ROUTE LENGTH	360.000	470.000	470.000	470.000
* 10	AVERAGE RIDING TIME	.640	.660	.770	.770
11	NUMBER OF TRANSFERS	.627	.627	.550	.550
12	AVERAGE TRANSFER DEL	3.400	3.200	1.890	1.867

PM = PERFORMANCE MEASURES

* = MORE IS BETTER,OTHERWISE LESS IS BETTER

( ) = DESIGN NO.

TABLE 6 TEST PROJECT 3: RANKING OF ALTERNATIVES, DESIGNS 100, 200, 300, AND 400

AVERAGE DOMINANCE RANKING((CONCORDANCE + DISCORDANCE)/2)

WEIGHTING SCHEMES	ALTERNATIVES 1	2	3	4
1	4.00(0)	2.50(0)	2.00(1)	1.50(1)
2	4.00(0)	2.50(0)	2.00(1)	1.50(1)
3	3.50(0)	2.00(0)	2.50(0)	2.00(1)
4	3.50(0)	2.00(0)	3.00(0)	1.50(1)
5	4.00(0)	3.00(0)	1.50(1)	1.50(1)
6	4.00(0)	2.50(0)	2.00(1)	1.50(1)
7	4.00(0)	2.00(0)	2.00(1)	2.00(1)
8	3.50(0)	2.00(0)	2.50(0)	2.00(1)
TOTAL	30.50	18.50	17.50	13.50

(1)--NON-DOMINATED; (0)--DOMINATED

FINAL RANKING OF ALTERNATIVES

RANK	ALT	DESIGN NO.
1	4	400
2	3(*)	300
3	2(*)	200
4	1(*)	100

(*)--THIS IS NOT A TOTALLY NON-DOMINATED ALTERNATIVE

The results of these three test problems clearly indicate that TNOP_ADVISOR is capable of providing advice that, when used to make design changes, will provide improved performance levels. Some evidence has shown that TNOP_ADVISOR can be applied to any design problem regardless of its base network, demand pattern, or route layout. Also it can be applied to the design of a new system or to the improvement of an existing system. In all three cases tested, TNOP_ADVISOR was able to provide the advice needed to find improved designs. The number of iterations it takes to find a high-performance design depends on the spatial complexity of the problem (size of network and

TABLE 7 SUMMARY OF TEST PROBLEM RESULTS

Test Problem	Avg. Dominance Ranking Start	Finish	No. of Iterations	Advice Generated
1	31.00	10.50	5	Add routes Change headways Change vehicle types Change layover time
2	24.00	12.50	14	Change headways Change vehicle types Change layover time Modify routes Change headways Change vehicle types Change layover time
3	31.00	10.50	5	Add routes Change headways Change vehicle types Change layover time

demand pattern), number of routes, number of vehicle types, policy headways on lines, objectives, and constraints of the problem undertaken. Also, making planning changes to the design will increase the number of iterations needed compared with making operational changes only.

Table 7 presents a summary of the results from all three tests. Shown are the gain made in each case in start to finish results, the number of iterations, and the advice given. This table indicates that in all three test problems the final design shows considerable and consistent improvement over the starting design. Test Problem 2 included planning changes, whereas the other two did not, so it required more iterations to reach the final design.

On average it takes 20 to 30 min to formulate a design (when network and demand input are ready) and simulate its performance with TNOP. For a new user, it may take much more time than this. To run PRISM and prepare the facts for the knowledge base takes about 10 min or more depending on the number of lines in the design. Running TNOP_ADVISOR takes another 15 min. The user needs a few minutes to a few hours to sift through the advice and prepare for the next iteration. All the times shown here are expected to increase with an increase in the size and complexity of the problem.

CONCLUSIONS

Conventional transit network design practice is user dependent and lacks consistency and reliability. It relies heavily on rules of thumb and principles developed by the user through study and practice over a period of years. But the methodology developed in this research, a KBES, captures the knowledge of an expert or experts and uses it to aid in the solution of this problem. The expert system method is much more systematic and consistent than traditional methods. However, some substantial human judgment and interpretive abilities are still required for the methodology developed in this study.

The interactive KBES approach was tested on three different test problems. Two of the test problems used the same base network and demand pattern, whereas the third one used a different base network and demand pattern. Two problems started from an initial stage and the third one started with a good existing design. TNOP_ADVISOR was able to provide advice that resulted in improvements to all three test designs, and each design created using the knowledge base was better than the preceding design. From the limited tests made, TNOP_ADVISOR has proved consistent and able to give appropriate and useful advice.

The interactive KBES methodology that uses TNOP to assist in the transit network design process has been successful. This methodology has many advantages over the traditional approach. Transit network design is an ideal application for a KBES. In the last decade, many expert systems have been developed in medicine, engineering, and other fields. Transportation planning has had few applications so far. From the results of this study, it is clear that transit network design is one area in which applications of expert systems should prove very useful.

One of the main advantages of this knowledge base approach is its transferability. The knowledge base developed in this study appears capable of giving appropriate advice irrespective of the demand pattern, network characteristics, or route layout. This is encouraging because many different problems can be assisted using the same knowledge base.

TNOP_ADVISOR provides capabilities to nonexpert users that allow them to generate transit network designs that are as good as those an expert user can generate.

REFERENCES

1. M. Rapp. Transit System Planning: A Man-Computer Interative Graphic Approach. In *Highway Research Record 415*, HRB, National Research Council, Washington, D.C., 1972, pp. 49–61.
2. Y. Chan. Configuring a Transportation Route Network via the Method of Successive Approximation. *Computers and Operations Research*, Vol. 1, 1974, pp. 385–420.
3. L. Silman, Z. Barzily, and U. Passy. Planning the Route System for Urban Buses. *Computers and Operations Research*, Vol. 1, 1974, pp. 201–211.
4. M. Rapp, P. Mattenberger, S. Piguet, and A. Robert-Grandpierre. Interactive Graphics System for Transit Route Optimization. In *Transportation Research Record 559*, TRB, National Research Council, Washington, D.C., 1976, pp. 73–88.
5. X. Rosello. An Heuristic Algorithm to Generate an Urban Bus Network. In *Advances in Operations Research: Proceedings of the 2nd European Congress on Operations Research*, Stockholm, Sweden, November 29–December 1, 1977 (M. Roubens, ed.). Elsevier, North-Holland Publishing Company, Amsterdam, 1977.
6. C. Mandl. Evaluation and Optimization of Urban Public Transportation Networks. *European Journal of Operational Research*, Vol. 5, 1980, pp. 396–404.

7. J. Lopez. *Operational Research '81* (J. Brans, ed.). Elsevier, North-Holland Publishing Company, Amsterdam, 1981, pp. 621–535.
8. T. Magnanti and R. Wong. Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, Vol. 18, No. 1, Feb. 1984.
9. R. Wong. Transportation Network Research: Algorithmic and Computational Questions. *Transportation Research*, Vol. 19A, 1985, pp. 436–438.
10. *Interactive Graphic Transit Network Optimization System (TNOP): User's Manual.* GM Transportation System Center, Nov. 1979.
11. C. Ramamoorthy, S. Shekhar, and V. Garg. Software Development Support for AI Programs. *Computer*, Jan. 1987, pp. 30–40.
12. L. Ortolano and C. Perman. A Planner's Introduction to Expert Systems. *American Planning Association Journal*, Winter 1987.
13. E. Shortliffe. *Computer-Based Medical Consultation: MYCIN.* American Elsevier, New York, 1976.
14. B. Buchanan, G. Sutherland, and E. Feigenbaum. Heuristic DENDRAL: A Program for Generating Explanatory Hypothesis in Organic Chemistry. *Machine Intelligence*, Vol. 4, American Elsevier, New York, 1969.
15. MACSYMA Reference Manual. Mathlab Group, Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, Mass., Nov. 1975.
16. R. Duda and J. Gasching. Knowledge-Based Expert Systems Come of Age. *Byte*, Vol. 6, No. 9, Sept. 1981, pp. 238–279.
17. D. A. Waterman. A Guide to Expert Systems. Addison-Wesley, Reading, Mass., 1985.
18. C. Yeh. Development of a Post-Disaster Transportation Network Trip Simulation Model and Traffic Recovery Strategy Expert System. Doctoral dissertation. University of Washington, Seattle, 1986.
19. N. Janarthanan and J. Schneider. Multicriteria Evaluation of Alternative Transit System Designs. In *Transportation Research Record 1064*, TRB, National Research Council, Washington, D.C., 1986, pp. 26–34.
20. P. Nijkamp and A. Van Delft. *Multicriteria Analysis and Regional Decision-Making.* Kluwer-Academic, Martinus Nijhoff, Norwell, Mass., 1977, 135 pp.
21. G. Giuliano, M. Berglund, K. Mortazavi, G. Thompson, P. Lyon, and J. Berechman. Making Transportation Corridor Investment Decisions within a Multimodal Framework. Final report. Institute of Transportation Studies, University of California, Irvine, Oct. 1983, 248 pp.

Expert Systems Development for Contingency Transportation Planning

ANSELMO OSVALDO BRAUN AND ANTONIO CARLOS MACHADO DE SÁ

Ensuring that the daily movement of millions of passengers and tons of goods in the Metropolitan Region of Rio de Janeiro—with a population of more than 10 million people—can be satisfactorily and safely accomplished is a complex task. It involves several transportation modes; federal, state, and local agencies; public and private operators; and skilled personnel. Very often parts of the system become ineffective. For many reasons, such as system component failures, strikes, accidents, disasters, or environmental catastrophes, special operational plans must be performed in order to recover service. In some of those agencies, plans are available for such conditions, but in most cases they are restricted to individual circumstances, mainly caused by equipment failures. Coordination among agencies and operators is almost nonexistent, but they work together when a critical situation arises. Another problem that has been addressed and that requires similar treatment is evacuation in the event of a nuclear disaster at the nuclear power plant in the state of Rio de Janeiro. For these numerous ill-structured problems, in which human behavior, social and political considerations, and multiobjective decision making are involved, the potential of expert systems technology was considered and is included in the contingency transportation planning of the Rio de Janeiro State Department of Transportation.

Presented in this paper is the research approach to contingency transportation planning taken by the Rio de Janeiro State Department of Transportation (SECTRAN/RJ).

The completion of operational plans for recovering transportation service in cases of disruptions caused by emergency or contingency conditions (restricted to a specific public or private operator) or joint plans (those involving more than one agency or operator) will be followed by the development of the corresponding expert systems to help in the decision-making process involved in that activity.

The inclusion of the expert systems technology in the project resulted from the 3-day Workshop on Expert Systems for Transportation, recently held in Rio de Janeiro, with the cooperation of the TRB Task Force on Expert Systems and ASCE Committee on Expert Systems. The workshop was organized by the Center for Transportation Technology (CETEC), an agency of SECTRAN/RJ, to disseminate within the transportation environment of Rio de Janeiro, especially among top administrators and transportation personnel at several federal, state, and local government agencies and private operators as well, information on the potential of expert systems applications in the transportation field. Besides the workshop, CETEC will address the development of expert systems for solving transportation problems, with emphasis on SECTRAN/RJ's contingency transportation planning research project, and will possibly focus on the development of special hardware for applications that may require such equipment. The first application, which is the subject of this paper, addresses the question of contingency transportation planning.

A. O. Braun, Empresa Brasileira de Planejamento de Transportes (GEIPOT), Rua Almirante Barroso 63, 29th floor, 20031, Rio de Janeiro, RJ, Brazil. A. C. Machado de Sá, Centro Tecnológico de Transportes (CETEC), Rua São José 35, 15th floor, 20010, Rio de Janeiro, RJ, Brazil.

PROBLEM IDENTIFICATION

The Metropolitan Region of Rio de Janeiro (RMRJ) is made up of several towns, with a resident population of more than 10 million people. Figure 1 shows a map of the transportation systems that operate in RMRJ. The modal split of passenger trips by public transportation service is shown in Table 1.

Besides the importance of service provided by the first three modes (commuter train, ferry, and subway), some lines or corridors carry more than 150 buses per hour per direction, in addition to the regular traffic of passenger cars and heavy trucks. Any interruption in the operation of the main components of the transportation network, especially during peak hours, leads to critical conditions, because there is no reserve of operational capacity in the system.

In the last 10 yr, some 15 strikes paralyzed the four modes mentioned above. In some cases, disruption of service lasted so long that the RMRJ was in total chaos. (The strike of commuter trains in February 1988, for example, lasted 11 days.) There were so many accidents and riots that people could not go to work or get back home.

As a result of the critical economic situation that the country has been experiencing in past years, social tension is a strong reason for authorities to worry about the potential of new events to cause operation of the transportation

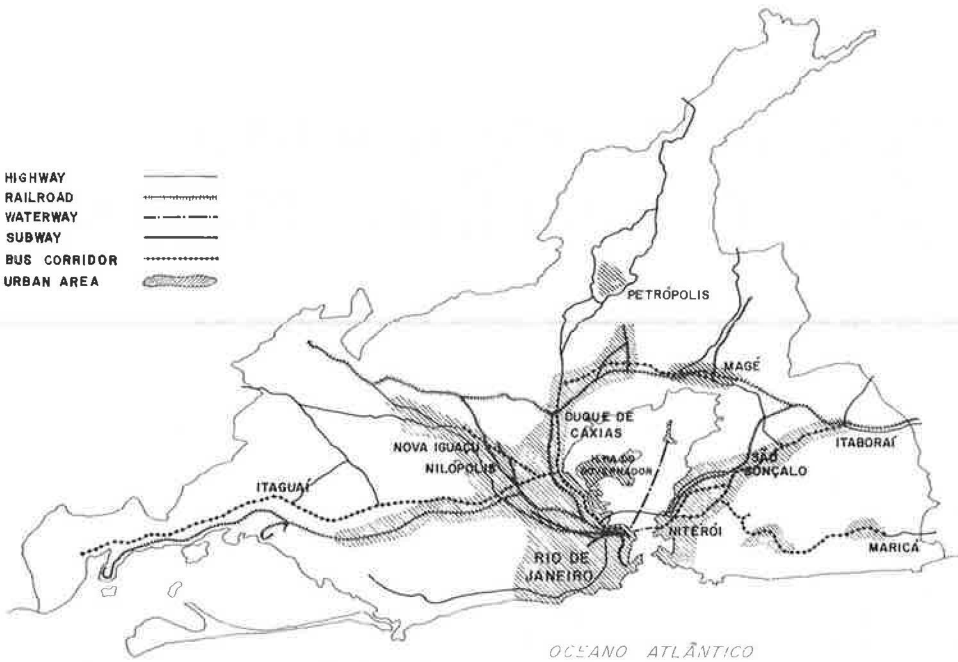

FIGURE 1 Transportation system in the Metropolitan Region of Rio de Janeiro.

TABLE 1 MODAL SPLIT OF TRANSIT IN RMRJ

Mode	No. of Lines	Passengers per Day No.	Passengers per Day Percent	Type of Operator
Commuter train	6	700,000	8.5	Federal government
Ferry	2	180,000	2.2	State government
Subway	2	300,000	3.7	State government
Local bus	850	7,000,000	85.6	Private
Total trips		8,180,000	100.0	

SOURCE: SECTRAN/RJ.

system to collapse. They are also afraid of serious injury to people and damage to the physical installation and rolling equipment. Records of such situations show that there have always been such undesirable consequences.

- In July 1988, a technical problem at a power station of the commuter train company caused disruption of service for 2½ hr, riots ensued, and five 12-car trains were damaged.
- The sliding of a barrier at the entrance to a highway tunnel in October 1987 and flooding in the suburban areas of Rio de Janeiro in January 1988 blocked traffic for 2 and 3 days, respectively, in the surrounding affected areas.
- Brazil started a nuclear power plant program years ago and built three plants at Angra dos Reis, in the state of Rio de Janeiro (Figure 2). The first plant suspended service 2 yr ago when a problem in a component of the reactor was detected.

These conditions were viewed as emergency situations for which special operational plans had to be prepared and performed.

In the event of an accident or nuclear disaster, a complex operation of transport mobilization must be carried out at once, in order to evacuate the population of towns and villages within the range of zones at risk and also allow for the necessary action to control and combat the problem.

It was also recognized that plans must be available for coping with such potential conditions and that strict cooperation between the agencies and operators involved must be exercised and implemented. After the Workshop on Expert Systems for Transportation, authorities agreed that expert systems constitute a convenient tool for the development and implementation of contingency transportation planning.

CONTINGENCY TRANSPORTATION PLANNING

SECTRAN/RJ (*1–5*) and its agencies (*6–10*), as well as non-transportation-related agencies (*11, 12*), have been working on the development of contingency transporta-

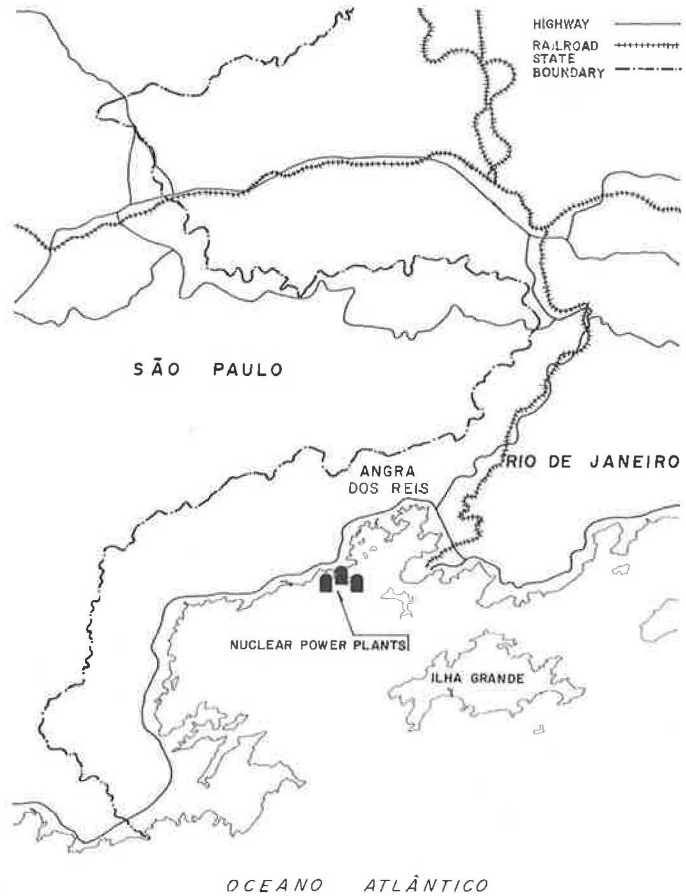

FIGURE 2 Location of nuclear power plants in Brazil.

tion plans in order to minimize the effects of service disruption and the inherent consequences of such conditions. However, it was recognized that without coordinating with other agencies of the federal and local government, as well as with private operators, inefficient action would result. The key word is preparedness, and this should involve all segments of the transportation system in the state of Rio de Janeiro, rather than agencies of the state of Rio de Janeiro.

To implement an alternative transportation operation for an emergency condition, it is necessary to count on the cooperation of other segments or components of the transportation system located in the affected area in the state.

This is exactly the critical issue of the problem being addressed: In addition to the conduct of research on individual problems of the several agencies with different statuses, structures, and objectives, the core problem to solve is how to address the institutional questions and the coordination of plans produced at those agencies.

Therefore, for expert systems development, the problem context comprises three main components: (a) contingency plans restricted to a specific transportation agency or operator, for emergencies that affect only a segment of the transportation system, the one for which a particular agency or operator is responsible; (b) joint contingency plans among agencies or operators, or both, for emergencies that affect more than one mode or segment of the transportation system in the RMRJ; and (c) a general contingency plan, which will be triggered at any agency or operator in a potential contingency condition but control of which will be centralized. Thus a first response of the plan is to recognize whether under the prevailing conditions there is in fact a contingency condition and, if so, whether it is restricted to only one or whether it also affects other system components.

Because of the complexity of the subject in the geographical area and from past experience and consequences that have resulted from such contingency conditions, the problem addressed constitutes effectively a challenging task for both strategists and knowledge engineering researchers. Furthermore, the issues involved in the research and development of a tool to help decision makers in such critical situations are particularly timely in the state of Rio de Janeiro.

The contribution of expert systems to the expected solutions should include not only reliable advice to operations managers and decision makers, because of the exemption from human weakness under critical circumstances that automated and well-tested systems present, but also a lasting methodology for problem solving under contingency conditions and a powerful tool for training managers in the various segments of the process both in individual mode agencies and at the central control agency.

Besides the expected benefits from the availability of contingency transportation plans, the research will induce managers from the individual agencies to organize and store their knowledge. It was found that in many of those agencies, very little had been written on preparedness and operational plans for contingency conditions, and the retirement of those experts meant that their knowledge was lost. However, some agencies developed very good manuals of procedures to cope with that situation, but again, most of them are restricted to cases of equipment failure.

The authors' particular interests at their corresponding agencies match those of SECTRAN/RJ, that is, structuring and supervising the development of a research methodology for the improvement of contingency transportation planning, which includes adoption of expert systems technology.

Nevertheless, even if adequate expert systems software will provide reasonable advice during consultation sessions for decision makers, make their work easier and more reliable, and reduce the risk of human failure, the applicability of such proposed expert systems strongly depends upon the completion of a preliminary phase of the research: the proper coordination among experts of the agencies and operators involved.

Fortunately, experts are available at most of those agencies. Despite the fact that there is usually one expert for a small part of the entire task at a particular agency or operator, which means that an extensive information collection task will be required, an in-house team at the central agency is organizing procedures for interviewing the experts, to prevent routine problems from being disregarded in the contingency plans.

RESEARCH PHASES AND CURRENT STATUS

As shown in Figure 3, the research is composed of three main phases.

Phase I has three activities. In Activity 1.1, basic criteria were developed at SECTRAN/RJ to identify what conditions should be of interest for contingency planning. In Activity 1.2, preliminary work was taken in order to identify what agencies should be involved in the contingency planning process and how to establish communication and coordination among them. It had already been determined that for many of the potential conditions that characterize a contingency situation and that may require mobilization of the transportation system, several non-transportation-related agencies must also be involved in the research and in the planning process as well. This caused considerable expansion of Activity 1.2, the current status of the research. In Activity 1.3, the basic criteria developed in Activity 1.1 will be reviewed and adjusted to the particular conditions of individual agencies. Then the alternatives of action that must be considered in any particular contingency condition will be selected. As the work progresses, at least two university research centers will be incorporated into the project. It is important to note that one set of procedures or actions will be considered for issues or problems restricted to each individual agency or operator and other sets for issues or problems related to one or more agencies or operators.

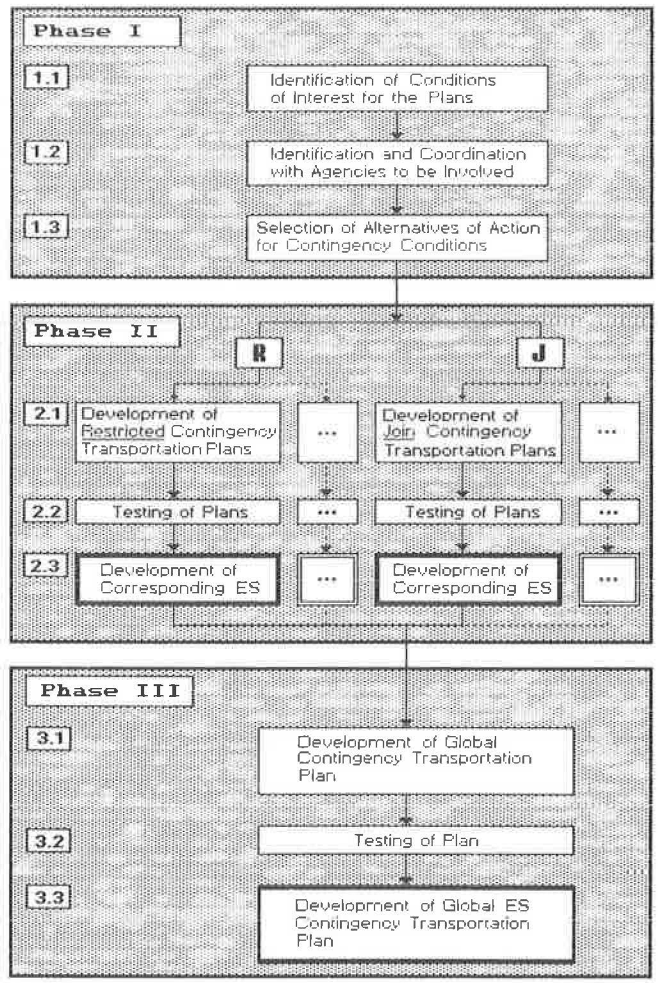

FIGURE 3 Phases of research.

Phase II has three types of activities. Activity 2.1 corresponds to the development of the Contingency Transportation Plan and Activity 2.2 to the testing of the plans. As a particular plan is tested, the corresponding expert system will be developed (Activity 2.3). However, there may be as many restricted plans and their corresponding expert systems, which individual agencies will have identified in Activity 1.3, as there are joint plans and their corresponding expert systems, which the research will also have identified in Activity 1.3.

Phase III differs from Phase II only in that rather than developing restricted or joint plans and expert systems, the coordination of those plans will be established and a Global Contingency Transportation Plan will be developed and tested, followed by development and testing of the Global Expert Systems Contingency Transportation Plan.

CONCLUSIONS

The current status of contingency transportation planning addressed by SECTRAN/RJ focuses on the replacement of individual action taken under contingency conditions by agencies and operators by an organized methodology and the contribution of the expert systems technology.

There are still many difficulties to cope with until the development of expert systems can take place in the process. A research approach, based on the key word "preparedness," was scheduled by SECTRAN/RJ, particularly at CETEC.

The research will lead managers and decision makers to contribute their expertise in the development of a lasting methodology for contingency planning by which their knowledge will be preserved. The final product of the research is likely to be very useful to the users and practitioners of transportation in operations, especially those who have to make decisions on how to recover service under the critical conditions of a contingency. The results will also serve for training new managers and decision makers in the field.

REFERENCES

1. Plano de Operações de Emergência na Área Nuclear—CNAAA-Unidade I. (Emergency Operation Plan for the Nuclear Zone-CNAAA-Unit 1.) SECTRAN/RJ, Rio de Janeiro, Brazil, 1987.
2. Plano de Emergência do Transporte Ferroviário. (Emergency Rail Transportation Plan.) SECTRAN/RJ, Rio de Janeiro, Brazil, 1986.
3. Plano de Emergência Intermunicipal. (Emergency Intermunicipal Plan.) SECTRAN/RJ, Rio de Janeiro, Brazil, 1986.
4. Plano de Emergência—Procedimento Operacional em Caso de Greve dos Rodoviários das Empresas Particulares. (Emergency Plan Operational Procedure in Case of a Highway Strike by Individual Firms.) SECTRAN/RJ, Rio de Janeiro, Brazil, 1985.
5. Ordem de Movimento—Plano de Evacuação na Região de Angra dos Reis. (Mobilization Method—Evacuation Plan for the Angra dos Reis Region.) SECTRAN/RJ, Rio de Janeiro, Brazil, July 1982.
6. Plano Emergencial de Apoio Rodoviário Visando a Evacuação da População na Área de Interferência da Usina Nuclear de Angra dos Reis. (Emergency Highway Aid Plan for the Evacuation of the Areas Affected by the Angra dos Reis Nuclear Power Plant.) Departamento de Estradas de Rodagem, Rio de Janeiro, Brazil, Feb. 1988.
7. Plano de Emergência—Bloqueios Rodoviários. (Emergency Plan—Highway Roadblocks.) Departamento de Estradas de Rodagem, Rio de Janeiro, Brazil, 1987.
8. Operação Carnaval de 1987. (Operations at the 1987 Carnaval.) Departamento Estadual de Trânsito do Rio de Janeiro, Rio de Janeiro, Brazil, 1987.
9. Operação Reveillon na Orla Marítima da Zona Sul do Rio de Janeiro. (Operation Reveillon on the South Shore of Rio de Janeiro.) Departamento Estadual de Trânsito do Rio de Janeiro, Rio de Janeiro, Brazil, 1986.
10. Plano Emergencial. (Emergency Plan.) Companhia de Transportes Coletivos, Rio de Janeiro, Brazil, Nov. 1983.
11. Usina de Angra—Medidas de Proteção à População. (Angra Nuclear Power Plant—Protective Measures.) Ministério do Interior, Defesa Civil, Brasilia, Brazil, 1987.
12. Usina de Angra—Como Você Pode Ajudar. (Angra Nuclear Power Plant—How You Can Help.) Ministério do Interior, Defesa Civil, Brasilia, Brazil, 1987.