REPRESENTING KNOWLEDGE FOR PORTFOLIO MANAGEMENT DECISION MAKING

Jae B. Lee
Management Science and Systems
School of Management
State University of New York at Buffalo

and

Edward A. Stohr
Computer Applications and Information Systems
Graduate School of Business Administration
New York University

September 1985

Center for Research on Information Systems
Computer Applications and Information Systems Area
Graduate School of Business Administration
New York University

Working Paper Series

CRIS #101

GBA #85-67

ABSTRACT

The business environment is complex and ill-defined. Broad areas of knowledge are needed to solve problems and each application area may require a different knowledge representation scheme. A system for portfolio investment advisors is being developed in Prolog at NYU which supports such an environment. It uses mixed knowledge representation schemes: production rules, logic, directed networks, and frames.

1. INTRODUCTION

Artificial intelligence (AI) techniques make it possible to achieve the objectives of DSS [7,13] since they can capture human intuition and judgement for supporting decisions involving semi-structured or unstructured tasks. This paper describes work at New York University on a prototype intelligent decision support system for portfolio management decision making (PMIDSS).

The PMIDSS interactively supports the decision process for an investment advisor interviewing a prospective investor. The role of the investment advisor is to provide data and advice to help individual investors construct/manage their portfolios effectively. The investment decision process can be regarded as a problem in decision-making under uncertainty [3]. It can be divided into two parts: (a) the analysis of investment timing, and (b) the investment selection decision (Table 1).

When compared to most scientifically oriented expert systems,

Domain	Problem Solving Phase	PS	KR Scheme
1. Investment Timing	Forecast stock market and match current situation to prototypes	1 2 3	Question Nets Networks Frames *Production Rules
2. Stock Selection	Select stocks from a candidate set.	4 5 6	Question Net Network of frames *Production Rules
* Production rules provide overall control for user interaction and reasoning			

Table 1: Correspondence of Problem-Solving Steps and KR Schemes

business expert systems encompass broader areas of knowledge where each area may require a different knowledge representation (KR) scheme [9]. The PMIDSS design reflects these requirements since it supports more than one problem domain each requiring a different architecture and reasoning.

Several other authors have described knowledge representations to achieve high system performance [5,8,11,14]. The need for more representational power has also led to experiments with combinations of knowledge representations (eg. AM [10] and CENTAUR [1]). In AM, domain knowledge for elementary set and number theory is represented in frames. Production rules which are attached to each slot in a concept frame are used to fill in the contents of a slot. CENTAUR uses similar representations to AM.

PMIDSS uses a mixed knowledge representation scheme: rules as Horn-clauses, logic, directed networks, and frames. We believe that greater flexibility and efficiency can be provided by these mixed representations than by a single-scheme based framework (eg. logic-based system). The different knowledge representation schemes correspond to different stages in the decision-making process. For example, frames are most suitable for classification while directed networks are useful for causal reasoning. This paper focuses on knowledge representation issues. Section 2 briefly reviews the knowledge requirements for portfolio management decision making and the system components. In Sections 3 and 4, it is shown how the logic programming language, Prolog, can be used as a basis for implementing mixed KR schemes. Section 5 presents some conclusions.

2. DOMAIN KNOWLEDGE REQUIREMENTS AND SYSTEM COMPONENTS

The problem of constructing a knowledge base for the portfolio manager is to uncover the rules and heuristics leading to a decision to purchase/sell particular securities. Once gathered, these rules must be suitably organized and encoded to support the reasoning process. We also allow certain 'expertise' to be exercised each time the system is run. This includes some subjective judgements about economic variables and the preference structure (or risk profile) of the investor.

The investment vehicles that will be included in the system include stocks, bonds, real estate and cash equivalents. Each requires specialized domain knowledge. Currently we are implementing only the stock-market portion of the investment decision. In PMIDSS, domain knowledge is divided into four different layers: general economy and stock market, industries, companies, and selected stocks. The knowledge in each layer consists of facts and heuristic procedures as well as rules. This expertise is used to forecast the stock market, to match the current economic situation to prototypical ones in the system, and finally to select the most promising stocks (see Table 1).

Several distinct reasoning systems use the domain knowledge to help solve the investment problem. These are: dialog management, control structure, model management, and data management. Dialog Management triggers some rules in the system and controls questions which lead to a particular decision. It supports user understanding, usability, and acceptance by coupling an explanation ability with the Control Structure and Knowledge Base (KB). The Control Structure (CS) consists of procedures to solve problems by reasoning. Since the expert knowledge in the KB cannot be used with complete certainty in reaching conclusions or recommending courses of action, CS interfaces with Model Management. This uses analytical tools (eg. mean-variance analysis [6]) to generate possible solutions at the request of the reasoning mechanism. Data Management supports the functions of PMIDSS by supplying the necessary data.

The two phases shown in Table 1, are each broken-down into three problem-solving steps:

Investment Timing:

1. Determine economic conditions
2. Select scenarios fitting economic conditions
3. Determine relative amount to invest in stocks, bonds etc.

Stock Selection:

4. Determine risk profile of the investor
5. Select particular stocks, bonds etc.
6. Determine amount to be invested in each selected asset

The system first determines general trends in the economy and stock market and analyzes the observed data to forecast future movements in the stock market. This step uses question nets (QN) for observation and a directed network representation which we call an expectation derivation net (EDN) for analysis. This employs causal reasoning [2]. A collection of frames is used for pattern-matching in step 2. A set of production rules represented as Horn-clauses

uses the output of the EDN and the frames to determine if the situation is favorable for common stock investment (step 3). Once the problem context to be solved has been determined, the system instantiates question nets to formulate investors' objectives (step 4). For step 5, a directed network of frames is used to select a set of promising stocks. This net is developed and explicitly stored during the problem session to facilitate explanation. With a set of promising stocks selected, a management science model is used to determine an optimal portfolio of stocks (step 6).

3. PROLOG AS A MIXED REPRESENTATION SCHEME

Prolog [4], implements a subset of first-order logic, the Horn-clauses. A Prolog clause has both a procedural and a declarative interpretation [12]. The three basic components in Prolog [9] are: terms which are either constants, variables, or structures and are used as arguments to predicates, predicates which can be used to name a specific relationship and form an assertion, and rules which are similar to production rules. These components allow implementation of a mixed representation scheme as shown below.

To define a predicate, more than one Horn-clause may be needed and a corresponding AND/OR graph is thus created for each such predicate. A Prolog-based knowledge base has a collection of these AND/OR graphs for rules and assertions. Since derivative events or facts can be represented in a directed network, the KB can have a collection of directed networks. It can also contain classified information in the form of frames. Directed networks and frames are composed of assertions. The KB can be modified through the introduction or deletion of Horn-clause so that the Horn-clause is an atomic unit.

As a vehicle for representing multiple schemes, Prolog provides several advantages. First, since it is a logic-based scheme, it provides a well understood formal semantics to verify the correctness of reasoning. Second, it prevents multiple uses of the same piece of knowledge by representing production rules, thereby inference power is enhanced. Third, since it can represent a network, it provides an efficient way to organize knowledge that must be directly retrieved. Fourth, it can provide a pattern matching ability by representing frames.

4. MODEL OF DOMAIN KNOWLEDGE IN PROLOG

We define a "portfolio management" (PM) concept as any assertion about the world that is used in the reasoning process (e.g. "interest rates will decrease"). In order to build a KB for PMIDSS and more effectively manipulate PM concepts, we need to develop knowledge representation conventions in the form of Prolog terms, predicates, and rules which will be meaningful to the user of system. This section provides some (much simplified) examples.

Rules. A rule can be represented in a Horn-clause with each clause using the assertions as well as the rules in the KB. An example in Prolog is as follows:

```
X:- Y, Z.          If Y and Z, then X.
X:- Y; Z.          If Y  or Z, then X.
```

X will be determined by the values of Y and/or Z. There may exist correlations between X and Y, and X and Z. These relations should be predetermined. The rules are used to represent procedural knowledge which can control user interaction and reasoning. Production rules provide the control steps of problem-solving and the rules of using AI data structures.

PM Concepts.

Concept	Sample Value	Prolog Assertion
stock price	increase, 50	increase(stock_price, 50).
interest rate	decrease, X	decrease(interest_rate, X).

The first assertion above is "the stock price is increased by 50 points". The values of stock price are subject to change and captured as variables such as X. As shown in the above examples, the concepts are represented as terms. The Prolog assertions are like formulae in Predicate Calculus.

Directed Network. A directed network consists of a set of nodes and links between them. Each node denotes a PM concept. A directed link means that a concept in the first node leads to another concept in the second node. Thus, the following network (called s-p-net):

```
A ----> B  ...  Y  ---->  Z

Legend: A, B, Y, Z: Nodes in the Network
```

can be represented in Prolog as follows:

```
-------------------------------------------
network (s_p_net, A, B).
network (s_p_net, B, C).
...
network (s_p_net, Y, Z).
is_network (s_p_net).
-------------------------------------------
```

The predicate "network" captures the PM concepts in the nodes. The first argument denotes the name of a specific directed network (in this case a causal net determining stock-price). The other two arguments in the "network" represent nodes linked by causation. Thus, concept A leads to B (eg. a decrease in interest rate leads to an increase in stock price). This was referred to above as an EDN. The interactive dialogue between a user and PMIDSS begins by using an EDN. Some of the PM nodes in the net are instantiated by answers to questions such as "Are higher interest rates predicted ?". Causal reasoning starts from these nodes to get to an intermediate and/or final node which contains a predicted result (e.g. the predicted change in stock prices).

<u>Frames of Assets</u>. Frames in PMIDSS are used to represent stereotypical situations. The following example represents a frame for favorable timing of investment in common stocks. A frame is composed of slots for variables and their values. Given a frame of cs-1,

```
      cs_1
     ________
X1 |___A____|
X2 |___B____|
.. | __  __ |
   |        |
Xj |___Q____|
```

the corresponding representation in Prolog is given below:

```
------------------------------------------------------------------
frame(cs_1, x1, A).
frame(cs_1, x2, B).
...
frame(cs_1, xj, Q).
is_frame(cs_1).
------------------------------------------------------------------
```

The first argument, "cs-1", denotes the name of the frame. In this case the frame represents phase 1 of the stock market cycle [9]. The second argument provides names for the slots. Here x1...xj (slots) are PM concepts that are useful in recognizing the different phases of the cycle. The third argument represents slot values. These are determined by interacting with the user via the EDNs. Fuzzy matching is used to select the frame best fitting the predicted economic conditions.

<u>Question Nets (QN)</u>. A QN consists of a set of disconnected subnets where a node represents a question to the user and an arc contains the user's answer which leads to another question and so on. Thus, a QN is simply a mechanism to allow prompt response sequences to be dependent on the answers obtained. For example, to analyze an economic condition, the system first activates the QN of economic analysis. This, in turn, may instantiate a PM node in an EDN. The causal reasoning mechanism then works on the EDN to analyze the situation.

<u>Directed Network of Frames</u>. A node in the net is a frame and an arc represents a rule or criterion to select the next similar frame. We can represent such a net in Prolog as follows:

```
-----------------------------------------------------
network(link_net, a, b).  /* Directed Network */
network(link_net, b, c).  /* of N Frames of CS */
...
network(link_net, m, n).
is_network(link_net).

frame(a, a_1, A).  /* Frame a */
frame(a, a_2, B).
...
frame(a, a_m, C).
...
frame(b, a_1, D).  /* Frame b */
frame(b, a_2, E).
...
frame(b, a_m, F).
...
frame(n, a_1, G).  /* Frame n */
frame(n, a_2, H).
...
frame(n, a_m, I).

is_frame(a).
is_frame(b).
...
is_frame(n).
-----------------------------------------------------
```

The predicate "network" denotes a directed network of (say) n nodes. Each frame has m slots. As an example, each node in the net may be the common stock of a company and a slot in the frame a price-earning ratio. The link between stock nodes may be done dynamically by stock selection rules such as "choose stocks that are in cyclically good industries for the current economy".

5. CONCLUDING REMARKS

In this paper we attempt to show how to formalize and organize large amounts of knowledge using a mixed KR scheme in Prolog. We believe that the KR schemes we have developed can be applied in similar application domains, thereby helping other system developers. EDNs and asset frames seem to be useful for forecasting and classification respectively. The directed network of frames seems to be useful for selecting where a large amount of knowledge is required but it is factorable using constraints.

We are currently implementing the reasoning programs of PMIDSS, and developing the knowledge by working closely with a number of professional financial analysts. We hope to demonstrate this work in later papers.

REFERENCES

[1] Aikins, J.S., Prototypical Knowledge for Expert Systems, Artificial Intelligence, 20, 1983.

[2] Asher, H. B., "Causal Modeling", Sage Publications, Inc., 1976.

[3] Clarkson, G.P.E., "A Model of the Trust Investment Process", in Computers and Thought, E.A. Feigenbaum and J. Feldman, eds., McGraw-Hill, 1963.

[4] Clocksin, W.F. and C.S. Mellish, Programming in Prolog, Springer-Verlag, 1981.

[5] Davis, R., B. Buchanan and E. Shortliffe, "Production Rules as a Representation for a Knowledge-Based Consultation Program", Artificial Intelligence, Vol. 8, No. 1, 1977.

[6] Elton, E.J. and M.J. Gruber, Modern Portfolio Theory and Investment Analysis, John Wiley and Sons, 1981.

[7] Ginzberg, M.J. and Stohr, E.A., "Decision Support Systems: Issues and Perspectives," in Decision Support System, M.J. Ginzberg, W.R. Reitman, E.A. Stohr, eds., North-Holland, 1982.

[8] Kowalski, R.A., Logic for Problem Solving, North-Holland, NY, 1979.

[9] Lee, J.B., An Intelligent Decision Support System for Portfolio Management Decision Making: Stock Selection, Ph.D Thesis, New York University, 1985, forthcoming.

[10] Lenat, D.B., "AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search", Rept. STAN-CS-76-570, Stanford University, 1976.

[11] Minsky, M., "A Framework for Representing Knowledge", The Psychology of Computer Vision, P.H. Winston, ed., mcGraw-Hill, 1975, pp211-277.

[12] Nau, D.S., "Expert Computer Systems", Computer, pp. 63-85, Feb., 1983.

[13] Pople, H.E. Jr., "Knowledge-Based Expert Systems: The Buy or Build Decision", in Artificial Intelligence Applications for Business, W. Reitman, ed., Ablex, 1984.

[14] Quillian, M.R. "Semantic Memory", in Semantic Information Processing, M. Minsky, ed., MIT Press, 1968.