TR/13/84 November 1984

O.R.,Statistics,A.I.- the potential for interdisciplinary progress

By

R.I. Phelps

O.R., Statistics, A.I. - the potential for interdisciplinary progress

R.I. Phelps

Brunel University

This paper examines the need for O.R. workers to become more involved in the development of A.I. A brief outline of A.I. is provided noting problems, techniques and objectives similar to those found in O.R. This outline gives an indication of how interdisciplinary development might proceed and indicates the direction in which O.R. training should be progressing.

Introduction

Considerable success has recently been achieved by the application of artificial intelligence (A.I.) work., in particular by the development of expert systems (1). In many areas this methodology competes with approaches based in O.R. or statistical methods and it is instructive to examine how it has achieved success in competition with these older and better established methods.

Its prime advantage may be seen as the ability to tackle complex problems by making use of subjective and heuristic methods similar to those used by humans. This enables processing of problems in a manner which may be suboptimal but which corresponds to human levels of performance and is therefore generally acceptable to clients whose main concern is for a workable and understandable system rather than an optimal one. Human reasoning has the ability to spot the essential elements in a problemand patterns in data, thus structuring the problem situation and allowing a qualitative analysis. It is relatively poor at handling quantitative, objective analyses.

O.R. and statistical approaches, on the other hand, attempt to build 'scientific' models which are quantitative in nature and emphasise the optimisation aspect of their techniques (2). This leads to problems when dealing with complex systems. Small parts can often be formulated as mathematical models, but the mathematical treatment of the whole problem is generally impossibly complicated if the modelling is to remain at all rigorous. Further, in such a 'scientific' framework it is difficult to include behavioural elements such as value judgements and reasoning by analogy. The result of this has been that O.R. and statistics have been most successful in dealing with self-contained technical problems within an overall problem scenario. Indeed, even in a case where the whole problem is essentially scientific and quantitative, medical diagnosis, it is interesting that expert system approaches (3) have had most success. Statistical approaches have been tried (4) but without great success since they have not attempted to model the symptom-disease process but have tried to directly correlate symptoms with diseases, and have found that to do so rigorously is not computationally possible. An O.R. modelling approach has not been attempted - perhaps because of the complexity of the problem or perhaps because this would go against O.R's perceived idea of itself as a decision aid rather than a decision maker. This is a situation where there is a strong case for a combination of the statistical, O.R. and A.I. approaches.

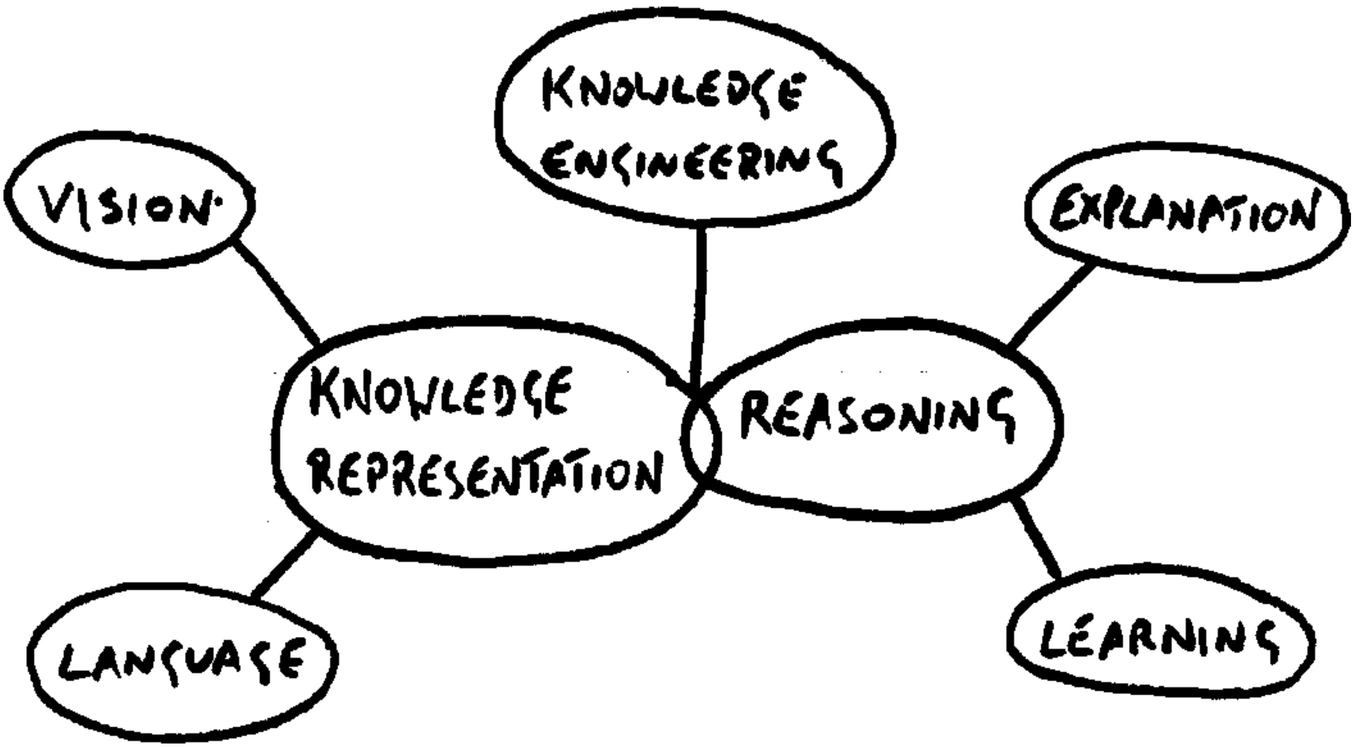
Another recent development is that of expert systems being constructed to carry out statistical work normally done by statistical specialists (5).

Continued ...

These trends point to a danger that O.R. and statistics nay find themselves overtaken by A.I. advances, ending up as no more than specialist suppliers of techniques which can be used within user-friendly knowledge-based packages, and that the real problem formulation and solution activities will be taken over by 'knowledge engineers'.

To avoid this it is necessary for O.R, and statistics workers to actively interest themselves in many aspects of A.I. There is already considerable interest in the use of expert systems as evidenced by the success of a number of conferences and meetings, e.g. (6). However, a narrow interest in the use of expert system packages and shells is not enough - expert systems in their present form may prove only a temporary showcase for A.I. as the underlying techniques are developed. It is necessary for the future of O.R. and statistics to understand the bases of A.I. as well as their applications. There are in fact many similarities between the subjects, both in the problems they face and the techniques they use: both the O.R. and A.I. approaches build models, both use 'heuristic' procedures in the absence of optimal ones, both are based in mathematics, both use computer implementations, both employ interdisciplinary teams. There are considerable areas of work in common where each side would benefit from a closer relationship with the other. Further, for the efficient solution of complex problems a combination of the approaches is clearly called for: objective models for those parts of a system capable of mathematical description together with human style heuristic reasoning for the more complex and behavioural parts. In order for this to be done it is necessary for O.R. and statistics workers to take an interest in A.I. as it relates to their own fields.

Continued ...



KNOWLEDGE ENGINEERING
VISION
EXPLANATION
KNOWLEDGE REPRESENTATION
REASONING
LANGUAGE
LEARNING

fig. 1

The main body of this paper sets out a brief outline of the main areas of A.I. and within each area points out problems that arise and solution methods that are used which are familiar in O.R. and statistics. A schematic outline of the areas to be covered is given in FIG.1.

Vision

Vision can be subdivided into three stages:

(1) Preprocessing

The input typically consists of an array of pixels each showing a level of grey (in black and white) or colour. The initial task is to locate lines (edges) and regions within the array in order to start picking out objects in the field of vision. Regarding a darker pixel as a higher value, this corresponds to finding peak regions and valleys separating them. Algorithms are used to find e.g. densest points, local peaks. Smoothing and relaxation (local consistency) techniques are employed (7). This work makes use of ideas also found in constrained and unconstrained optimisation methods, clustering algorithms and dynamic programming, all widely used in statistics and O.R.

(2) 3-D Information

There are two basic approaches to obtaining information aboutthe 3-D structure of a scene once the lines and regions have been located, a)intrinsic information given by local properties of the array, e.g. reflectance,shadows, stereo pair matching, b) knowledge-driven

Continued ...

expectations about the scene, such as knowledge of the kind of objects likely to be present and their properties, e.g. the feasible ways in which lines can connect the corners of a polygon to each other. This knowledge is used to cut down the possible number of 3-D interpretations of the scene and to hypothesise the presence of certain 3-D objects (8).

(3) Recognition

Once the scene has been analysed into separate 3-D shapes, the next question is whether they can be recognised as known objects. There are two approaches to this, a) Template matching: known objects are stored as one or more templates encoding their shapes as visual wholes. The input image is matched to the templates and 'recognised' if a close match is found. E.g.

TEMPLATE : E

INPUTS : E εE

Dynamic programming can be used to match shapes where natural variation consists of non-linear expansions and compressions relative to the template (9). b) Feature extraction: known objects are stored as a set of features, and input shapes presented as a set of input features. Matching is between the sets of features rather than whole objects, The substages of this process involving matching input features against template features are themselves template matches.

OBJECT : E

FEATURES : $\Gamma \succ$ L A x 90°

The relevant question is which features of an object should be used. Given a large

Continued ...

set of possible features dynamic programming can be used to find the most efficient set within cost constraints. The number of possible features can be reduced by statistical techniques such as principal components or factor analysis (10). Partitioning objects into classes with known numbers can be achieved using linear programming (11).

Language

Language processing can be considered in two parts.

(1) Speech recognition.

In order to have a linguistic input it is firstly necessary to recognise which words are being spoken (written input requires visual recognition discussed under vision), The approaches used adopt two main ideas. a) Knowledge-driven expectations to limit the number of words searched for matching at any one time, on the basis of context, b) Matching template waveforms for syllables and word with input waveforms:

It is necessary to allow for speed of utterance, accent, emphasis, etc., involving non-linear distortions and so is conveniently formulated using dynamic programming (12).

(2) Language understanding.

Two major areas are contained here, a) Syntactic analysis of the form of language i.e. analysis and application of the rules of grammar for a language, b) Semantic analysis achieved by encoding the meaning of the input in a specially designed conceptual representation language, typically consisting of a limited number of fundamental concepts which are sufficient to encode the meanings expressed in natural language (13) . Although these studies can involve mathematical models such as grammar structure, this area has little immediate interface with statistics and 0.R.

Knowledge Engineering

This is the process of obtaining knowledge and heuristic lines of reasoning from an expert. The first problem is the decision to use one expert or several. If the latter, then there are problems of combining group preferences and expert judgements. These problems have been studied at some length in the decision analysis literature , e .g. (14).

Given an expert, the problem of obtaining information from him is again well-known in O.R. Problems of different responses in hypothetical situations to those occurring in the real world have been studied in gaming (15) and the problem of obtaining information without imposing unnatural structure on it is a starting point for 'soft' methodologies such as cognitive mapping (16).

Once a description of the expert's approach is obtained, there remains the problem of analysing it for structure and content. Humans may not be able to give good explanations of how they make certain decisions. In these cases, techniques such as multidimensional scaling (17), familiar in statistics and O.R., can be used to discover underlying reasons for these decisions.

Knowledge Representation and Reasoning

It has been proposed to represent knowledge in three main ways:

(1) Analogue.

A mental model isomorphic to the real world is formed. The best example of this is pictorial imagery; it is suggested that mental images are viewed to obtain information in the same way that real scenes are viewed. Thus the visual

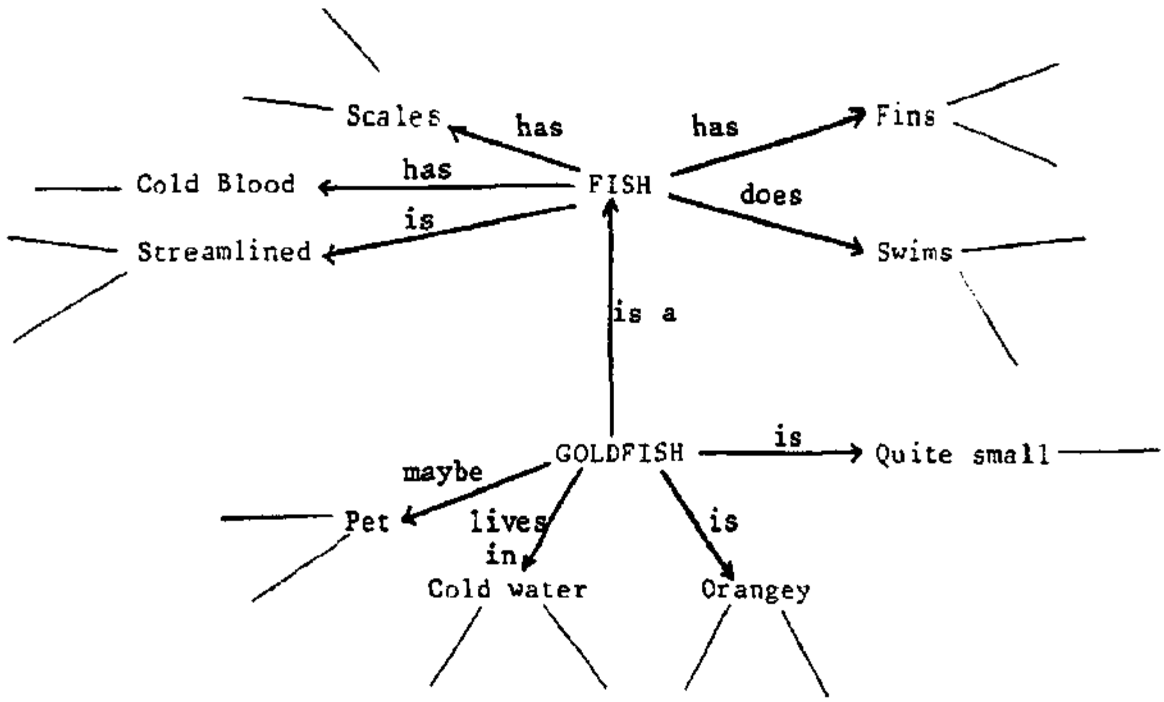
Continued ...

knowledge is represented in an essentially pictorial (analogue) manner.

Reasoning associated with this form of representation takes the form of inspecting or manipulating the analogue. Thus, to decide whether or not the bisector of angle of a triangle bisects the opposite side, one or more bisected triangles might be mentally viewed and the effect of their bisectors noted (18).

2) Associative.

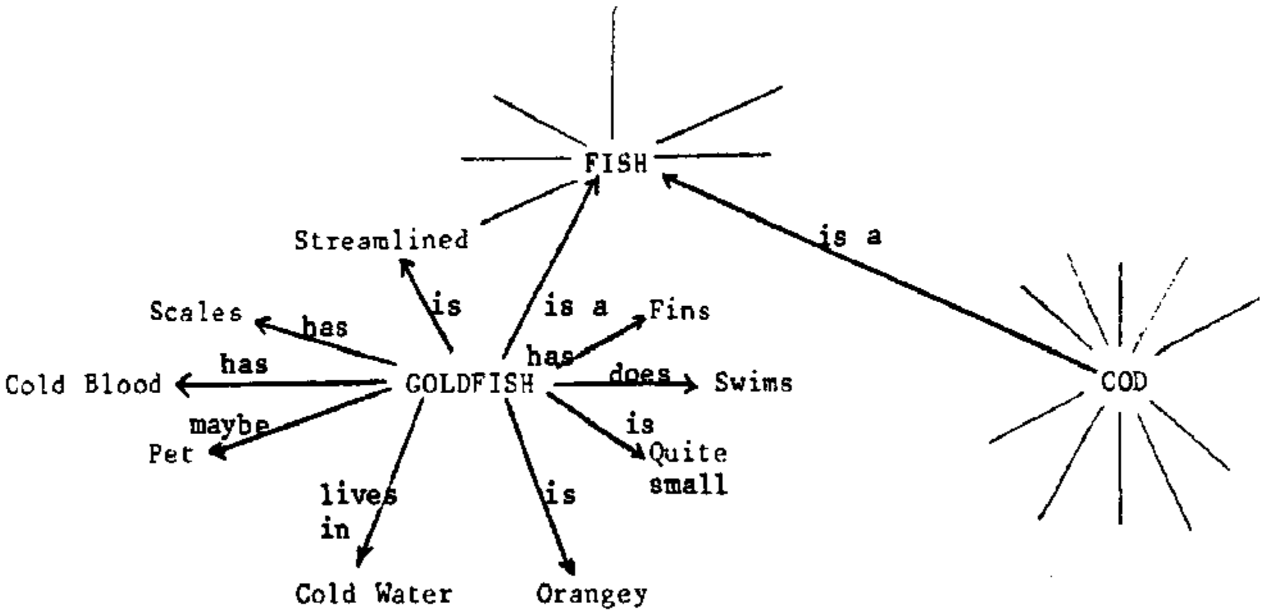
A network is formed whose node consist of concepts, objects and features, and whose arcs denote relationships between nodes, A part of such a network might look like:



This is a network utilising hierarchical category relations, where subordinate concepts (Goldfish) inherit the properties of higher order concepts (Fish), so that e.g. 'has fins' does not need to be represented again at Goldfish level (19). Alternatively the network could be organized on a prototypical basis where each concept is stored with all of its 'typical' features and higher order concepts are

Continued ...

stored with the concepts typical of them.



Reasoning using the network structure takes the form of activation of nodes spreading through the network via the arcs until a sought after node is activated (or search is terminated). Relevant techniques for reasoning in a network are therefore search procedures well known in O.R. including shortest-route algorithms (A*) and the branch and bound heristic ($\alpha - \beta$ pruning).

(3) Propositional.

Facts are coded in propostitional form e.g. 'A goldfish is a fish', 'A goldfish has fins'. Propositional reasoning is carried out by using inference rules e.g. 'If X is a creature with fins and a tail then X is a fish'. Systems using only rules in the 'If.. .then...' format are the special case of production systems; another special case is logical representation (20). The rules of logic can be applied to logical statements to deduce valid conclusions, but for all but very small systems heuristic rules are necessary to guide the search for combinations of propositions which give useful deductions. Further, many statements do not fall naturally into a logical form. The production system approach of independent declarative rules has the advantage of manageability - facts or rules can be added or deleted independent of the rest of the knowledge base, but the disadvantage of needing more control rules to decide how the inference rules should be applied to input data and the knowledge base. This is the form used by the majority of existing expert systems. The problem of which inference rule to apply next can be approached both by knowledge-driven heuristics and by sequential statistical decision theory.

Combinations of these forms of knowledge representation are possible. The concept of 'frame' representations essentially consists of grouping together knowledge into useful chunks. A frarae for Goldfish might look like:

<u>Goldfish</u>

KINDOF	:	FISH
COLOUR	:	ORANGE
SHAPE	:	STREAMLINED
HABITAT	:	COLD WATER
PARTS	:	BODY, HEAD, FINS,TAIL
SIZE	:	QUITE SMALL

There would also be frames for other concepts, e.g. Fish, streamlined, fins, etc. Thus the frames are connected with-each other via the attributes held in their 'slots'. They therefore form a network-like structure. The links between the frames are represented by propositional procedures attached to the frames,' e.g. for the Goldfish frame, a procedure might be: 'For information on sight, go to the Head frame'. Frames combine network, and propostional ideas in knowledge representation (21). In addition, it is not necessary to pair one form of knowledge representation with the same form of reasoning as outlined in 1,2,3. In particular, the combination of network representation with prepositional reasoning has been adopted e.g. (22).

Problems common to all these schemes are:

(a) Focusing: because of processing limitations it is only efficient to 'focus' upon and use a small subset of the available knowledge base at any one time (23). The control mechanism for determining the optimal subset is generally knowledge driven, i.e. problem specific heuristics, but sequential statistical methods are also relevant.

(b) For ward and backward chaining: should reasoning proceed forward from the existing state to attempt to reach a goal, or backward from a goal to find whether it can be reached from the present state (24)?. Heuristic search procedures as previously mentioned are of use here.

(c) Treatment of uncertainty: as in O.R. the two most prevalent formalisms for handling uncertainty are probability and fuzzy sets, but there is also considerable use of ad-hoc measures . Fuzzy set theory is sometimes associated with prototypical knowledge representation (25).

(d) Multiple objectives; in multiobjective situations the various goals have to be evaluated as a whole. The approaches developed in O.R. and statistics for handling these problems will become more relevant as more difficult applications are tackled (26).

Continued ...

Explanation

For user acceptance of an expert system it is necessary that the system be able to explain its reasoning. Present attempts to do this consist largely of regurgitating the sequence of rules that were successfully applied in order to reach the solution (27). However, this tends to produce a large list incorporating many minor steps and checks, and does not resemble a human explanation. There is a problem of how major steps in the reasoning process can be picked out to form a human style explanation. This is essentially the same problem facing an O.R. or statistical analyst making a presentation to his client. A compact, understandable explanation of the results must be devised taking into account the user's needs and background. This is an important topic in both areas.

Learning

Learning can be achieved on two levels. The more superficial level is the changing of parameter values within a given procedure. No attempt is made to find any useful structure in the system's input. This level is often called adaptive learning. The second and deeper level does attempt to find structure by initially classifying the input into categories and then deciding responses on the basis of the categories constructed. It is the problem of categorisation that is fundamental to learning. Learning can also be taught from examples or be untaught. In taught learning examples of input from different classes are presented and the system uses these known examples to form discrimination procedures for these classes.

Continued ...

(1) Taught Learning.

a) Adaptive: adjusting parameter values in tlie light of incoming information e.g. adjusting numerical values which evaluate the 'goodness' of positions in board games (28). This problem is also studied in Bayesian statistics and control theory.

b) Partitioning methods: rules are devised to partition input into a given number of categories. This can be done logically or statistically. In both cases the attributes of the input which can be used for discrimination must be given to the system. In the logical case, rules in predicate logic are constructed sufficient to partition the examples into their correct classes (these rules may be expressed as a discrimination tree) e.g.

'If length < 10 and temperature < 60 and colour = orange then Goldfish.'

The rules so formed can be generalised to deal with a wider class of input than the training examples, e.g. the above rule might be generalised to:

'If length < 10 and colour = orange or red then Goldfish.' (29)

Statistically, discriminant analysis can be used. This is more powerful than the logical method in dealing with 'noise' in the input and is more efficient, but offers less 'transparent' discrimination rules which are therefore difficult to generalise,

(2) Untaught Learning.

In the absence of examples known to be from different classes, the problem is to find 'natural' classes or patterns in the input. This can be approached from several angles. Operators may be provided which can detect certain sorts of patterns; hierarchical 'conceptual' clusters can be found in abstract descriptions (30). Perceptual clusters can be sought indicating structure in the input (30). Little work has been done on this topic of central importance to A.I., but in this area exploratory data analysis methods such as cluster analysis are of obvious importance.

Valididation

A final problem faced both in O.R. and A.I. is that of model validity. Expert systems, like O.R. models, can be 'fine-tuned' by adjusting internal parameters to perform well in a specific situation. It is difficult to know whether the methods involved will generalise to other situations, i.e. whether the basic approach and structure of the model is sound. This problem is exacerbated in knowledge-based systems where the basic idea is to use problem specific knowledge and heuristics. This is a problem requiring further work in both fields.

Conclusions

The similarity of problems and methods arising in many aspects of O.R., Statistics and A.I. has been discussed. It is hoped that an awareness of the existence of such similarities will give rise to greater cross-disciplinary involvement in the underlying principles of A.I. as well as in the application of developed techniques such as expert systems. In many cases essentially the same problem is being tackled by these disciplines, each in its own way, sometimes using similar approaches sometimes different ones. There is a need for workers in each field to become more aware of what is happening in the others. In particular there is a need to bring together the data exploration and decision-theoretic techniques of statistics, the problem formulation, mathematical programming and optimisation techniques of O.R. and the expert systems approach of A.I. Such a synthesis would provide the tools necessary to evaluate alternative approaches to the same problem and the ability to integrate these tools into a powerful form of decision aid.. This is an area in which O.R. courses should be providing training for the future.

References

1 The Organisation of Expert Systems A tutorial. M.Stefik et al. Artificial Intelligence 18, 135-173. 1982.

2 Decision Methodology. D.J.White. Wiley.

3 Artificial Intelligence in Medicine. P.Szolovits . Colorado: Westview Press. 1982.

4 Evaluation of clinical decision aids, with an application to a system for dyspepsia. D.J. Spiegelhalter. Statistics in medicine 2, 207-215. 1983.

5 An expert system for regression analysis. W.A. Gale & D. Pregibon. In: Computer Science and Statistics: Proceedings of the 14th Symposium on the interface. Springer Verlag. 1982.

6 Computer Assisted Decision Making. London, April 1984.

7 A survey on image segmentation K.S. Fu & J.K. Mui. Pattern Recognition 13, 3-16. 1981.

8 Computer Vision Systems. A.R. Hanson & E.M. Riseman. Academic Press. 1978.

9 On-line recognition of hand-written characters utilising postional and stroke vector sequences. K. Ikeda et al. Pattern Recognition 13, 191-206. 1981.

Continued...

10 Patterns in pattern recognition 1968 — 1974 - L. Kanal. IEEE Trans. Inf. Theory 20, 697-722, 1974.

11 Network Flows, Transportation and image processing. M.K.S. Tso. O.R. Conference, Lancaster, U.K. 1984.

12 Automatic Speech Recognition using local timescale variability information M.J. Russell et al. Proc. Inst. Acoustics. November 1982.

13 Scripts, Plans, Goals and understanding. R.C. Schank and R.P. Abelson. Lawrence Erlbaum. 1977.

14 An axiomatic approach to expert resolution. P.A. Morris. Mgmt. Sci. 29, 24-32. 1983.

15 Human behaviour in games. J.H. Klein. Young O.R. Conference, U.K. 1984.

16 Thinking in organisations. C. Eden et al. Macmillan. 1979.

17 Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis. J.B. Kruskal. Psychometrika 29, 1-26. 1964.

18 Imagery and internal representation. S.M. Kosslyn. In Cognition and Categorisation Ed. E. Rosch & B.B. Lloyd, Erlbaum. 1978.

19 Explorations in cognition. D.A. Norman and D.E. Rumelhart. Freeman. 1975.

Continued . . .

20 Principles of artificial intelligence. N.J. Nilsson. Tioga Press. 1980.

21 A framework for representing knowledge. M. Minsky.
In The psychology of computer vision. Ed. P. Winston. McGraw-Hill. 1975.

22 Semantic network representations in rule-based inference systems.
R.O. Duda et al. In Pattern directed inference systems. Ed. D.A. Waterman & F. Hayes-Roth. Academic Press. 1978.

23 An overview of production systems. R. Davis & J.J. King.
In Machine Intelligence 8. Ed. E. Elcoch & D. Michie. Ellis Horwood. 1977.

24 Planning techniques for rule selection in deductive question answering.
P. Klahr. In Pattern directed inference systems. Ed. D.A. Waterman &
F. Hayes-Roth. Academic Press. 1978.

25 Models of concepts. B. Cohen & G.L. Murphy. Cognitive Science 8, 27-58. 1984.

26 Conflicting objectives in decisions. D.E. Bell, R.L. Keeney, H. Raiffa.
Wiley. 1977.

27 Handbook of A.I. Vo1.2. Ed. A. Barr & E.A. Feigenbaum. Pitman. 1982.

28 Some studies in machine learning using the game of checkers. A.L. Samuel.
I.B.M. J. Research & Development 3. 210-220. 1959.

Continued ...

29 Machine learning. Ed. R.S. Michalski et al. Tioga Press. 1983

30 A cluster model of learning. R.I. Phelps & P.B. Musgrove. O.R. Society conference, Lancaster, U,K. 1984.