

**THE BCS PROFESSIONAL EXAMINATION
Professional Graduate Diploma**

April 2003

EXAMINERS' REPORT

Knowledge based Systems

Question 1

Discuss two commercial areas where artificial intelligence techniques are currently being used successfully to improve business performance. Your answer should include a description of the techniques being used, and a discussion of how they improve performance. (25 marks)

Answer Pointers

A number of areas could be discussed. Here are some examples:

- Drug discovery. Finding relationships within genetic data is difficult using standard statistical techniques. Genetic algorithms, for example, are good at finding relationships where there are a large number of variables, but a small number of cases. This research is facilitated by the rich availability of AI and statistical techniques and the ability of modern computer programs that implement the techniques for the researcher on the selected data set.
- Direct mail. Marketing departments in many of the larger commercial companies now use decision tree and neural network technology to identify people who are most likely to respond to marketing campaigns. The application of AI algorithms in this environment is interesting in that the end result can save a company thousands of pounds each year – because they send mail to fewer people and yet maintaining a similar number of responders. Therefore the company has similar revenue expectations from a lower initial spend. The people applying these algorithms often have very little knowledge of how the algorithm works – they are more interested in testing the algorithm for correctness and the subsequent application of the model.
- Fraud detection. Fraud departments use algorithms such as decision trees to identify possible fraudulent transactions. As well as identifying possible fraud, it is important to keep the number of false positives to a minimum. The benefit to the company is large both in terms of actual money saved from identifying fraud early and in terms of customer satisfaction.

Question 2

Outline the main components of Natural Language Processing (NLP). (10 marks)

Stating any assumptions made, discuss how NLP could be used to:

- Gain an understanding of the relationships between electronic documents contained within a library (5 marks)
- Automatically archive electronic documents (5 marks)
- Translate Language (5 marks)

Answer Pointers

Part (a): Book work.

Part (b):

- By gaining an understanding of the content of each document, it is possible to relate the content of different documents to each other. With the provision of a suitable interface, these relationships can be seen visually.
- By producing a taxonomy and understanding the similarities of documents within contained in each category, new documents can be archived automatically, by gaining an understanding of their content, and placing the document in the most compatible category.
- This involves having an understanding of the text and being able to relate this to the grammatical structure of a new language.

Question 3

Outline the mechanism by which rule induction works. How does this mechanism differ from a decision tree? (15 marks)

Identify and describe an algorithm for converting a decision tree into a set of rules. (10 marks)

Answer Pointers

Rule induction is an example of a covering algorithm. Such algorithms operate by adding tests to a rule that is under construction, always trying to create a rule with maximum accuracy. A set of rules is learned one at a time. At each step:

1. Find a rule that covers a large number of positive examples within the test data set without covering any negatives.
2. Remove the positive examples it covers.

In comparison, a decision tree algorithm iteratively chooses an attribute to maximise the separation between the classes. For example:

1. For each attribute, compute its entropy with respect to the conclusion
2. Select the attribute with lowest entropy.
3. Build a tree with branches:
 - if $A=a_1$ then ... (subtree1)
 - if $A=a_2$ then ... (subtree2)
 - ...etc.

Where A is the selected attribute and a_1, a_2 etc. are the values A can take.

4. For each subtree, repeat this process from step 1.
5. At each iteration, one attribute gets removed from consideration. The process stops when there are no attributes left to consider, or when all the data being considered in a subtree have the same value for the conclusion.

It is possible to devise a decision tree to replace a set of rules, automatically. A decision tree has the form

```
If attribute1 = value1 then <subtree 1>
else if attribute1 = value2 then <subtree 2>
else if ...
.....
else if attribute1 = valueN then <subtree N>
```

This corresponds directly to a set of rules, with as many rules as there are leaf nodes in the tree. Each rule is a tracing out of the path from the top of the tree to a leaf node.

Question 4

- a) What are the requirements of a learning system? (5 marks)
- b) Describe the generic principles behind the machine learning process. (10 marks)
- c) Explain with reasons why machine learning is important in building expert systems. (4 marks)
- d) Describe **two** major categories of learning algorithms with their advantages and disadvantages. (6 marks)

Answer Pointers

- a)
- bottleneck
 - desirable in problems that lack algorithmic solutions
 - ill defined or informally stated problems
- b)
- determining a description of a given concept from a set of concept examples
 - concept examples can be positive or negatives
 - representation and languages used to describe the examples and concepts
 - possible values of the attributes and their hierarchies
 - learning algorithm builds on the type of examples, on the size and relevance of the background knowledge, on the representation issues, on the presumed nature of the concept to be acquired
- c)
- able to deal with imperfections of data
 - incomplete data, missing attributes
 - errors, noise
- d)
- black box methods, e.g. neural networks or mathematical statistics
 - black box approaches develop their own concept representation that is used for concept recognition purposes. This internal concept representation cannot be easily interpreted by the user and provides neither insights nor explanation of the recognition process
 - knowledge oriented methods
 - aim to creating symbolic knowledge structure that satisfy the principle of comprehensibility

From Machine Learning and Data Mining by Ryszard Michalski, Ivan Bratko and Miroslav Kubat

Question 5

You have been asked to advise a company on how to develop and manage an expert system.

- a) Produce a checklist of 10 items for planning and managing the development. (10 marks)
- b) Produce a checklist of 10 items on the selection of domain. (10 marks)
- c) Explain how you would evaluate the success of the expert system. (5 Marks)

Answer Pointers

(a)

- set well defined project objectives, and make sure all people concerned understand them
- use well qualified staff
- in projecting milestones and making promises, be conservative without losing support
- view selection of the expert system domain as an important phase
- view the selection of the development environment as an important phase
- plan to develop, relatively rapidly, a feasibility demonstration system with limited scope
- make clear all concerned about the limitation of the feasibility prototype system
- develop full prototype system
- test and evaluate
- for the production system, explore the possibility of changing the system with a view toward what is best for the deployed system
- documentation for operators, maintainers and users
- determine mode of deployment
- marketing and pricing
- training
- system maintenance group for both knowledge and programme

(b)

(10 marks)

- where conventional programming is not likely to be effective
- there should be recognised experts who solve the problems
- a need to capture the expertise
- it should be expected that the system may not produce optimal or correct results 100% of the time
- completed system should be expected to have a significant payoff
- primarily symbolic reasoning
- use of heuristics
- should not require significant use of common sense reasoning
- task is clearly defined
- all inputs available to the experts should also be available to the system
- commitment from the experts
- amount of knowledge should be large enough to make the knowledge based developed nontrivial
- close working relationship between developers and domain experts
- supported by top management
- test cases are readily available
- domain should be fairly stable
- payoff should be measurable

(c)

- measure the competence of the system by the degree of its agreement with the known correct results
- evaluate the system by comparison against human experts
- When the domains allows, utilise for system evaluation experts not associated with the project
- Test the system in the field under actual operational condition
- Set standards of evaluation
- Verify that the implemented expert knowledge contains no internal errors, such as redundant rules, sets of circular rules, illegal slot values. Utilise automated checking systems if available.

From Developing and Managing Expert Systems by David Prerau.