IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2009

MEng Honours Degree in Information Systems Engineering Part IV
MSci Honours Degree in Mathematics and Computer Science Part IV
BSc Honours Degree in Mathematics and Computer Science Part III
MEng Honours Degrees in Computing Part IV
MSc in Advanced Computing
MSc in Computing Science (Specialist)
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the Associateship of the City and Guilds of London Institute
This paper is also taken for the relevant examinations for the Associateship of the Royal College of Science

PAPER C493=I4.48=E4.41

INTELLIGENT DATA AND PROBABILISTIC INFERENCE

Wednesday 13 May 2009, 14:30
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1 Figure 1 shows a Bayesian belief network. Each variable modelled by the network has just two states which are written in lower case, node A having states a_0 and a_1.

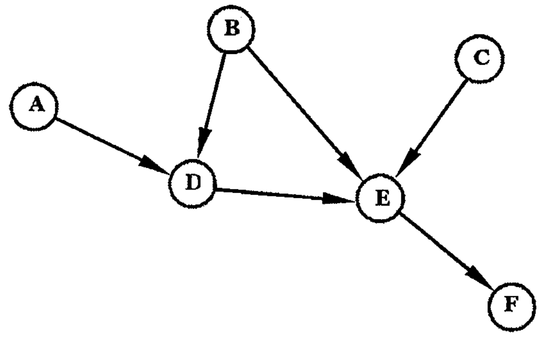

Fig. 1: A Bayesian belief network

The prior probabilities of the root nodes are as follows:

$P(A) = [0.8, 0.2]$, $P(B) = [0.5.0.5]$, $P(C) = [0.5, 0.5]$

The conditional probability tables are as follows:

	a_0b_0	a_0b_1	a_1b_0	a_1b_1
d_0	0.5	0.5	0.8	0.4
d_1	0.5	0.5	0.2	0.6

	$b_0c_0d_0$	$b_0c_0d_1$	$b_0c_1d_0$	$b_0c_1d_1$	$b_1c_0d_0$	$b_1c_0d_1$	$b_1c_1d_0$	$b_1c_1d_1$
e_0	0.5	0.5	0.9	0.7	0.5	0.5	0.7	0.9
e_1	0.5	0.5	0.1	0.3	0.5	0.5	0.3	0.1

	e_0	e_1
f_0	0.8	0.3
f_1	0.2	0.7

Pearl's operating equations for probability propagation are given in a supplementary sheet.

a Calculate the π evidence for node F after the network initialisation has occurred.

b Given that C is instantiated to state c_1, A is instantiated to a_1 and D is instantiated to d_1 calculate the λ evidence for node B.

c Why is propagation not possible in the case where B, D and E are not instantiated, but F is instantiated?

d Explain how a hidden node could be used to transform the above network into one in which probability propagation can always be carried out. Re-draw the network showing how the hidden node is connected.

e It turns out that the sensor from which the instantiated values of node B are calculated becomes stuck so that B is always in state b_0. Accordingly the network is simplified by the removal of node B. Calculate the conditional probability matrices $P(D|A)$ and $P(E|D\&C)$ in the simplified network.

The five parts carry, respectively, 25%, 20%, 15%, 20%, and 20% of the marks.

2 The network defined in question 1 is to be transformed into a join tree for the purpose of probability propagation.

a Draw the moral graph corresponding to the network of question 1, and indicate and label the cliques in it.

b i) Set up a table of the cliques showing their W R and S variable sets, and the potential function attached to each clique.

ii) Explain what is meant by the running intersection property, and draw a join tree of your cliques that observes the running intersection property.

c Explain briefly how the λ message from a clique to its parents can be calculated from the potential table associated with that clique. Given that F is instantiated to f_0 and there are no other instantiations, find the λ message from the clique where F is in the R variable set.

d Explain how the λ message calculated in part c is used to update the parent node of the clique from which it is sent. Calculate the updated potential table for the parent node in this case.

e Explain, with reference to your join tree, how π messages are sent. (There is no need to calculate the actual values for this tree and instantiation)

The five parts carry, respectively, 20%, 25%, 15%, 25%, and 15% of the marks.

3 a Explain what is meant by the Markov blanket of a node in a Bayesian network. For the network shown in figure 1 what are the Markov blankets of nodes D and F?

b Given that, in figure 1, the nodes A, C and F are instantiated explain how a Monte Carlo Markov Chain (MCMC) method could be used to find the probability distributions over the uninstantiated nodes (B, D and E).

c Compare and contrast the use of the MCMC methods with the join tree algorithm for making inferences in Bayesian networks. Consider cases where the variables are all highly dependent, and there are therefore many arcs in the network, and those where there is little dependency between the variables, and therefore there are few arcs in the network.

d Explain briefly how the re-sampling technique called alternatively “Arcing (adaptive resampling and combining)” or “Boosting” is used in constructing Bayesian network classifiers.

The four parts carry equal marks.

4a The Mahalanobis distance is defined by the following equation:

$$M(\mathbf{X}_1, \mathbf{X}_2) = \sqrt{(\mathbf{X}_1 - \mathbf{X}_2)\Sigma^{-1}(\mathbf{X}_1 - \mathbf{X}_2)^T}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are points in a multidimensional space and Σ is the covariance matrix of a particular class.

i) Explain with the aid of a suitable diagram why the Mahalanobis distance is an important measure in allocating points to classes.

ii) Is it necessary to use it in cases where the covariance matrix is diagonal?

b A face recognition system has a data base of fifty subjects, each stored as an image of resolution 128 by 256. The images are all packed into a large data matrix (D) of dimension 50 by 32768, in which each row is a complete subject image. Explain, in terms of matrix operations, how the covariance matrix (Σ) of this data can be calculated.

c Principal component analysis is carried out on this covariance matrix. At best, how many eigenvectors with real non zero eigenvalues can be found from this matrix? Explain why they are called "eigenfaces" in the context of face recognition.

d Explain how, and why, the eigenvectors of the covariance matrix are used in face recognition.

e Given a set of eigenfaces explain how an image of a real face may be reconstructed.

f The covariance matrix used above is of dimension 32768 by 32768, and consequently calculating its eigenvectors is slow and prone to numerical errors. Describe fully a technique for calculating the same real non-zero eigenvectors from a much smaller matrix (50 by 50 in this case).

The six parts carry, respectively, 15%, 15%, 15%, 15%, 15%, and 25% of the marks.

Supplementary information: Pearl's Operating Equations

Equation 1: The λ message

$$\lambda_c(a_k) = \sum_{i=1}^{n} \pi_c(b_i) \sum_{j=1}^{m} P(c_j|a_k\&b_i)\lambda(c_j)$$

and for the case of the single parent there is a simpler matrix form:

$$\lambda_{\mathbf{c}}(\mathbf{A}) = \lambda(\mathbf{C})\mathbf{P}(\mathbf{C}|\mathbf{A})$$

Equation 2: The π Message from A to C is given by:

$$\pi_C(a_j) = \begin{cases} 1 & \text{if } A \text{ is instantiated for } a_j \\ 0 & \text{if } A \text{ is instantiated but not for } a_j \\ P'(a_j)/\lambda_c(a_j) & \text{if } A \text{ is not instantiated} \end{cases}$$

Equation 3: The λ evidence of node C with n children $D_1, D_2, ..D_n$:

$$\lambda(c_j) = \begin{cases} 1 & \text{if } C \text{ is instantiated for } c_j \\ 0 & \text{if } C \text{ is instantiated but not for } c_j \\ \prod_i \lambda_{D_i}(C_j) & \text{if } C \text{ is not instantiated} \end{cases}$$

Equation 4: The π evidence of node C with two parents A and B:

$$\pi(c_i) = \sum_{j=1}^{n} \sum_{k=1}^{m} P(c_i|a_j\&b_k)\pi_c(a_j)\pi_c(b_k)$$

This can be written in matrix form using as follows:

$$\pi(\mathbf{C}) = \mathbf{P}(\mathbf{C}|\mathbf{A}\&\mathbf{B})\pi_{\mathbf{C}}(\mathbf{A}\&\mathbf{B})$$

where

$$\pi_C(a_j\&b_k) = \pi_C(a_j)\pi_C(b_k)$$

Equation 5: The posterior probability of variable C:

$$P'(c_i) = \alpha\lambda(c_i)\pi(c_i)$$

where α is chosen to make $\sum_i P'(c_i) = 1$