Intelligent Systems: Reasoning and Recognition

James L. Crowley

ENSIMAG 2 / MoSIG M1 Practice Exam

Conditions: You have the right to use any notes or written material. You may answer questions in English or in French. When appropriate, illustrate your answer with mathematics. Your written answers must be clear and legible. Illegible text will not be graded. Duration: 3 hours.

1) (4 points) Write a critical evaluation of the technology for expert systems. What can be done with these techniques? What are the limitations?

What is an expert system? An "Expert System" is a computer program that emulates the decision making ability of a human expert. Expert systems are designed to solve complex problems by reasoning with a symbolic expression of knowledge that has been provided by one or more domain experts.

Expert Systems are typically organized with at least two key components:
1) A hard-coded symbolic expression of the reasoning process of a domain expert, and
2) A generic "inference engine".

The expert system solves to problems posed by a user by using the inference engine to reasoning with the domain knowledge.

What can be done with an expert system?
Expert systems technologies can be used to build problem solving systems for domains where problems can be solved using automated reasoning procedures with a large number of unstructured "facts". Such domains include medical diagnosis, mechanical diagnosis, logistics planning, systems configuration and process monitoring.

Limitations:
Expert systems suffer from a number limitations.
1) They use superficial or "syntaxique" reasoning and do not include any deeper understanding of the problem or its solutions.
2) Such systems are generally unable to detect when they have been asked to solve a problem outside of their domain of expertise. In such cases they can provide non-sense answers.
3) The construction of the knowledge base is a tedious and expensive process requiring extensive interaction and testing with a domain expert.
4) Such systems can not adapt or learn from their mistakes.

Grade Scale:

Some statement, nothing false	1
Vague answer, general idea	2
Partially Correct, incomplete or rambling	3
Clear, concise, complete answer	4

(half points allowed)

2) (4 points) Assume the following class definitions in CLIPS:

```
(defclass PERSON (is-a USER) (role abstract)
     (slot NAME (create-accessor read-write) (type STRING))
     (slot ID (create-accessor read-write))
     (slot father (create-accessor read-write) (default unknown))
     (slot mother (create-accessor read-write) (default unknown))
     (multislot brothers (create-accessor read-write) (default unknown))
     (multislot sisters (create-accessor read-write) (default unknown))
)

(defclass MAN (is-a PERSON)
     (role concrete)(pattern-match reactive)
     (slot wife (create-accessor  read-write) (default unknown))
     (slot gender (storage shared)(default male) (create-accessor read))
)

(defclass WOMAN (is-a PERSON)
     (role concrete)(pattern-match reactive)
     (slot husband (create-accessor  read-write) (default unknown))
     (slot gender (storage shared) (default female)
          (create-accessor read))
)
```

a) Create a set of rules to build a family structure by asking for the wife for each man, the husband for each woman, and the father and mother for each person. Names these rules (ask-wife, ask-husband, ask-father, ask-mother).

```
(defrule ask-wife
     ?M <- (object (is-a MAN) (NAME ?n) (ID ?HID) (wife  unknown))
=>
     (printout t "Who is the wife of " ?n "? ")
     (bind ?NAME (readline))
     (bind ?ID (sym-cat ?NAME))
     (send ?M put-wife ?ID)
     (if (neq ?ID nil) then
          (make-instance ?ID of WOMAN (ID ?ID) (NAME ?NAME)
               (husband ?HID)))
)

(defrule ask-husband
     ?M <- (object (is-a WOMAN) (ID ?n) (husband  unknown))
=>
     (printout t "Who is the husband of " ?n "? ")
     (bind ?ID (read))
     (send ?M put-husband ?ID)
     (make-instance ?ID of MAN (ID ?ID) (wife ?n))
)

(defrule ask-father
     ?M <- (object (is-a PERSON) (ID ?n) (father  nil))
=>
     (printout t "Who is the father of " ?n "? ")
     (bind ?ID (read))
     (send ?M put-father ?ID)
     (make-instance ?ID of MAN (ID ?ID))
```

```
)

(defrule ask-mother
     ?M <- (object (is-a PERSON) (ID ?n) (father  nil))
=>
     (printout t "Who is the mother of " ?n "? ")
     (bind ?ID (read))
     (send ?M put-father ?ID)
     (make-instance ?ID of MAN (ID ?ID))
)
```

b) Define a set of message handlers for the class PERSON that can determine the names of the paternal and maternal grandparents.

```
(defmessage-handler PERSON paternal-grandfather ()
     (bind ?g-father (send ?self:father get-father))
 (send ?g-father get-ID)
)

(defmessage-handler PERSON paternal-grandmother ()
     (bind ?g-mother (send ?self:father get-mother))
 (send ?g-mother get-ID)
)


(defmessage-handler PERSON maternal-grandfather ()
     (bind ?g-father (send ?self:mother get-father))
 (send ?g-father get-ID)
)

(defmessage-handler PERSON paternal-grandmother ()
     (bind ?g-mother (send ?self:mother get-mother))
 (send ?g-mother get-ID)
)
```

Grade Scale:

2 points for each correctly completed part

-0.5 point for minor syntax errors, -1 point for partially incorrect rule

3) (4 points) Assume the following temporal relations between intervals A, B, C and D.

Event A before Event B :	(A < B)
Event B during event D :	(B d D)
Event A starts event C:	(A s C)
Event D overlaps event C :	(D o C)

a) What are the possible relations of A to D obtained by transitivity with B?

T(A < B, B d D) = { >, o, m, d, s}

b) What are the possible relations of A to D obtained by transitivity with C?

T(A s C, C oi D) = {oi, d, f}

c) What are the possible relations of A to D after constraint propagation?

(A d D)

Grade Scale: 1 point for questions a and b. 2 points for question c. .

4) (8 points) You have been hired as a political analyst. You are working on the political campagne for a referendum. Your job is to identify the sectors of the population for which you can design targeted publicity. For this you prepare a questionnaire for a poll. Each question has a small number of possible response. The questions are as follows:

1) What is your age? A) 18-29, B) 30-39, C) 40-49, D) 50-59, E) 60 or older
2) What is your sex? A) Male, B) Female.
3) What is level of education? A) High-school B) University Bachelor, C) Masters Degree D) Doctorate, E) Other.
4) What is your annual Salary? A) < 15 000 B) 15 001 to 30 000 C) 30 001 to 60 000 D) 60 001 to 90 000, E) More than 90 000.
5) How will you vote in the referendum? A) Yes, B) No, C) Undecided, D) I do not plan to vote.

a) (2 points) For the group who have responded A or B in Question 5, explain how to use a ratio of histograms to predict the most likely vote for each category of age. How many persons should be polled? How can you determine the probability of error for your prediction?

let Age_m represent the age of person *m* with possible values of {A, B, C, D, E}
Let $Vote_m$ represent the vote of person *m* with possible answers of {yes, no, undecided, abstain}

Define three hashs h(Age), h_{yes}(Age) and h_{no}(Age) each with 5 possible values {A, B, C, D, E}.

Given the responses of a group of M persons {A_m} and {V_m}, count the responses.

∀m=1, M: IF V_m=yes THEN $h_{yes}(Age_m) = h_{yes}(Age_m) + 1$ ELSE
IF V_m=no THEN $h_{no}(Age_m) = h_{no}(Age_m) + 1$

∀age: $h(age) = h_{yes}(A_{ge}) + h_{no}(A_{ge})$

Bayes rule says:

$$p(Vote = yes \mid Age = A) = \frac{p(Age = A \mid Vote = yes)}{p(Age = A)} p(Vote = yes)$$

thus:

$$p(Vote = yes \mid Age) = \frac{\frac{1}{M_{yes}} h_{yes}(Age)}{\frac{1}{M} h(Age)} \frac{M_{yes}}{M} = \frac{h_{yes}(Age)}{h(Age)}$$

With 5 categories of Age, you need to poll at least 5 x 10 = 50 persons who vote yes, and 5 x 10 = 50 who vote no.

To determine the probability of error, you must sum the fraction of the histogram of the class that is NOT maximum over all values of age.

Let the most likely vote be $\hat{V} = \arg\max_{Vote} \{ p(Vote \mid Age) \}$

IF $\hat{V} = V$ THEN Correct ELSE Error

$$p(Error \mid Age) = 1 - p(\hat{V} \mid Age) = 1 - \frac{h_{\hat{V}}(Age)}{h(Age)}$$

$$p(Error) = \sum_{age} p(Error \mid Age) = \sum_{age} 1 - p(\hat{V} \mid Age) = \sum_{age} (1 - \frac{h_{\hat{V}}(Age)}{h(Age)})$$

b) (2 points) Explain how to use a ratio of histograms to predict the response to question 5 as a function of the answers to questions 1, 2, 3, and 4. How many people must you poll? How can you predict the probability of error?

Let $\vec{X} = \begin{pmatrix} Age \\ Sex \\ Education \\ Salary \end{pmatrix}$ be a response with Age $\in$ {A, B, C, D, E}, Sex $\in$ {A, B}, Education $\in$ {A, B, C, D, E}, Salary $\in$ {A, B, C, D, E}.

Then as before create 5 histograms, for each possible response to Vote V $\in$ {A, B, C, D}

$\forall$m=1, M: h($\vec{X}_m$) = h($\vec{X}_m$) + 1
IF V_m=A THEN $h_A(\vec{X}_m)$ = $h_A(\vec{X}_m)$ + 1 ELSE
IF V_m=B THEN $h_B(\vec{X}_m)$ = $h_B(\vec{X}_m)$ + 1 ELSE
IF V_m=C THEN $h_C(\vec{X}_m)$ = $h_C(\vec{X}_m)$ + 1 ELSE
IF V_m=D THEN $h_D(\vec{X}_m)$ = $h_D(\vec{X}_m)$ + 1

$$p(V \mid \vec{X}) = \frac{h_V(\vec{X})}{h(\vec{X})}$$

How many people must you sample?

Number of Cells = 5 x 2 x 5 x 5 = 250
Need 10 x 250 = 2500 responses for each answer to question 5.
(8 x 10 is ok).

Probability of Error ?

Let the most likely vote be $\hat{V} = \arg\text{-}\max_{Vote}\left\{p(Vote \mid \vec{X})\right\}$

$$\text{IF } \hat{V} = V \text{ THEN Correct ELSE Error}$$

$$p(Error \mid \vec{X}) = 1 - p(\hat{V} \mid \vec{X}) = 1 - \frac{h_{\hat{V}}(\vec{X})}{h(\vec{X})}$$

$$p(Error) = \sum_{age} p(Error \mid \vec{X}) = \sum_{\vec{X}} (1 - p(\hat{V} \mid \vec{X})) = \sum_{\vec{X}} (1 - \frac{h_{\hat{V}}(\vec{X})}{h(\vec{X})})$$

c) (2 points) Explain how to use a quadratic discrimination formula to predict the response to question 5 given the responses to questions 1, 2, 3, and 4.

Answer:
A quadratic discrimination function requires computing the mean and variance for the features. This requires that the feature values take on numerical values, rather than symbolic values.

For example, for age, education, and salary, this can be integers {1, 2, 3, 4, 5} rather than symbols {A, B, C, D, E}. this is possible because these answers have an order relation.

The question about sex poses a potential problem. There is no order relation between the genders M and F, so computing an average is not meaningful. None the less, the values 1 and 2 can be used for the answers for this question because there are only two possible values.

A convenient way to do this is compute the histogram for the training date as in part b). Here, A, B, C and D represent the possible responses to question 5.

∀m=1, M: $h(\vec{X}_m) = h(\vec{X}_m) + 1$
IF V_m=A THEN $h_A(\vec{X}_m) = h_A(\vec{X}_m) + 1$ ELSE
IF V_m=B THEN $h_B(\vec{X}_m) = h_B(\vec{X}_m) + 1$ ELSE
IF V_m=C THEN $h_C(\vec{X}_m) = h_C(\vec{X}_m) + 1$ ELSE
IF V_m=D THEN $h_D(\vec{X}_m) = h_D(\vec{X}_m) + 1$

Then compute the moments for the histograms. Let v ∈ {A, B, C, D} represent the possible answers, and $h_v()$ be the histogram for each answer k.

$$p(\vec{X}) = \mathcal{N}(\vec{X};\vec{\mu},\, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}\det(\Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu})^T\, \Sigma^{-1}(\vec{X}-\vec{\mu})}$$

and

$$p(\vec{X} \mid v) = \mathcal{N}(\vec{X};\vec{\mu}_v,\, \Sigma_v) = \frac{1}{(2\pi)^{\frac{D}{2}}\det(\Sigma_v)^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu}_v)^T\, \Sigma_v^{-1}(\vec{X}-\vec{\mu}_v)}$$

$$\text{where: } \vec{\mu}_v = E\{\vec{X}_m^v\} = \frac{1}{M_v}\sum_{m=1}^{M_k} \vec{X}_m^v = \frac{1}{M_v}\sum_{\vec{X}} h_v(\vec{X})\cdot \vec{X}$$

and

$$\Sigma_v = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \dots & \sigma_{1D}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \dots & \sigma_{2D}^2 \\ \dots & \dots & \dots & \dots \\ \sigma_{D1}^2 & \sigma_{D2}^2 & \dots & \sigma_{DD}^2 \end{pmatrix} \quad \text{where: } \Sigma_v = E\{(\vec{X}_m^v - E\{\vec{X}_m^v\})(\vec{X}_m^v - E\{\vec{X}_m^v\})^T\}$$

$$\sigma_{ij}^2 = E\{(X_{im}^k - \mu_k)(X_{jm}^v - \mu_v)\} = \frac{1}{M_k}\sum_{m=1}^{M_v}(X_{im}^v - \mu_v)(X_{jm}^v - \mu_v)$$

you need to create 4 discriminant functions, one for each possible vote

$$\vec{g}_v(\vec{X}) = \begin{pmatrix} g_A(\vec{X}) \\ g_B(\vec{X}) \\ g_C(\vec{X}) \\ g_D(\vec{X}) \end{pmatrix}$$

where $g_v(\vec{X}) = -\frac{1}{2} Log\{\det(\Sigma_v)\} - \frac{1}{2}(\vec{X} - \vec{\mu}_v)^T \Sigma_v^{-1} (\vec{X} - \vec{\mu}_v) + Log\{p(v)\}$

This can be written as a quadratic polynomial

$$\boxed{g_v(\vec{X}) = \vec{X}^T D_v \vec{X} + \vec{W}_v^T \vec{X} + b_v}$$

where: $D_v = -\frac{1}{2}\Sigma_v^{-1}$

$\vec{W}_v = -2\Sigma_v^{-1}\vec{\mu}_v$

and $b_v = -\frac{1}{2}\vec{\mu}_v{}^T \Sigma_v^{-1}\vec{\mu}_v - Log\{\det(\Sigma_v)\} + Log\{p(v)\}$

d) (2 points) Explain how to use the EM algorithm to discover categories of voters who are likely to vote non given their responses to questions 1, 2, 3, and 4.

We can represent the set of responses for each vote, v, as a sum of N Gaussians. Each Gaussian represents a category of voters. We can discover the categories for evaluating performing the EM for increasing values of N. We stop increasing N when the Nth Gaussian has a very small sum of a_n parameters. This is equivalent to saying that the category has very few voters.

EM is an iterative two-step process in which we alternately estimate $\{\vec{\mu}_n, \Sigma_n\}$ and $\{a_n\}$. This is performed by an iterative algorithm known as
EM: Expectation Maximisation

The EM algorithms constructs a table, h(m, n)

h(m, n) = P{the voter m is from category n}

Choose N (the number of categories).
set *i=1*.
Form an initial estimate for $\vec{v}^{(1)} = (\alpha_n^1, \vec{\mu}_n^1, \Sigma_n^1)$ *for n = 1 to N.*

For example, set the initial values as $\alpha_n^1 = \frac{1}{N}$, $\vec{\mu}_n^1 = n\vec{\mu}_0^1$, $\Sigma_n^1 = I$
or with any reasonable first estimation. The closer the initial estimate, the faster the algorithm converges.

Expectation step (E)

Calculate the table $h(m,n)^{(i)}$ using the training data and estimated parameters.

$$h(m,n)^{(i)} = p((h_m = n) \mid \{X_m\}, \vec{v}^{(i)})$$

$$h(m,n)^{(i)} = \frac{\alpha_n \mathcal{N}(\vec{X}_m, \vec{\mu}_n, \Sigma_n)}{\sum_{j=1}^{N} \alpha_j \mathcal{N}(\vec{X}_m, \vec{\mu}_j, \Sigma_j)}$$

Maximization Step (M)

Estimate the parameters $\vec{v}^{(i+1)}$ using $h(m,n)^{(i)}$

M: (Maximisation)

$$S_n^{(i+1)} := \sum_{m=1}^{M} h(m,n)^{(i)}$$

$$\alpha_n(i+1) := \frac{1}{M} S_n(i+1)$$

$$\mu_n(i+1) := \frac{1}{S_n(i+1)} \sum_{m=1}^{M} h(m, n)^{(i)} X_m$$

$$\Sigma_n^{(i+1)} := \frac{1}{S_n^{(i+1)}} \sum_{m=1}^{M} h(m,n)^{(i+1)} (\vec{X} - \vec{\mu}_n^{(i+1)})(\vec{X} - \vec{\mu}_n^{(i+1)})^T$$

The Log-likelihood of the parameter vector is

$$Q^{(i)} = \ln\{p(\{\vec{X}_m\} \mid \vec{v}^{(i)})\} = \sum_{m=1}^{M} \ln\left\{\sum_{j=1}^{N} \alpha_j^{(i)} \mathcal{N}(\vec{X}_m \mid \mu_j^{(i)}, \Sigma_j^{(i)})\right\}$$

It can be shown that, for EM, the log likelihood will converge to a stable maximum. The change in Q will monotonically decrease. When

$\Delta Q = Q^{(i)} - Q^{(i-1)}$ is less than a threshold, halt.