CAIRO UNIVERSITY

Institute of Statistical Studies and Researches

Course: CS605 + IS605 (Advanced Topics) | Date: 10th January, 2012
Instructors: Dr. Khaled Wassif | Final Exam | Allowed Time: THREE Hours

Attempt ALL Questions:

Q1) What is a production system model? Describe the basic components of an expert system. (10 Marks)

Q2) What is a conflict set of rules? How can we resolve a conflict? Describe the basic conflict resolution methods (12 Marks)

Q3) What is uncertainty? What is the Bayesian rule? Write the Bayesian rule in terms of hypotheses and evidence. Give an example of the rule representation in the expert system based on Bayesian reasoning. (12 Marks)

Q4) Suppose an expert; given four conditionally independent evidences E_1, E_2, E_3 and E_4, creates three mutually exclusive and exhaustive hypotheses H_1, H_2 and H_3. The following table illustrates the prior and conditional probabilities provided by the expert:

	Hypothesis		
Probability	i = 1	i = 2	i = 3
$p(H_i)$	0.40	0.35	0.25
$p(E_1\|H_i)$	0.3	0.7	0.4
$p(E_2\|H_i)$	0.8	0.0	0.5
$p(E_3\|H_i)$	0.5	0.5	0.6
$p(E_4\|H_i)$	0.2	0.3	0.7

If the evidences are observed in the order E_4, E_2, E_3 and then E_1; compute the posterior probabilities for all hypotheses. (20 Marks)

Q5) What are the likelihood of sufficiency and likelihood of necessity? How does an expert determine values for both likelihoods? (12 Marks)

Q6) Define a linguistic variable and its value. Give an example. How are linguistic variables used in fuzzy rules? Give three different examples of fuzzy rules. (12 Marks)

Q7) Define a certainty factor. How does an expert system establish the net certainty for conjunctive and disjunctive rules? Give an example for each case. (12 Marks)

Q8) How are objects related in frame-based systems? What are the 'a-kind-of' and 'a-part-of' relationships? Give examples. (10 Marks)

With My Best Wishes

CAIRO UNIVERSITY
INSTITUTE OF STATISTICAL
STUDIES & RESEARCH
MAY 2015
TIME: 3 HOURS

COMPUTER SCIENCE DEPARTMENT
MASTER OF COMPUTER SCIENCES
ADVANCED TOPICS IN COMPUTER
SCIENCE (CS605)
DR. HESHAM A. HEFNY

ANSWER THE FOLLOWING QUESTIONS

Q1: a) Prove that the S-norm and the T-norm operators are related by the following relation: $S(x, y) = 1 - T(1-x, 1-y)$. (12 points)

b) If the S-norm operator is selected to be: $\frac{a+b-(2-g)ab}{1-(1-g)ab}$, where $g \in [0,\infty]$.

i- compute the corresponding T-norm operator.

ii- Write down the T-norm and the corresponding S-norm for the cases: g=0,1,2.

Q2: A fuzzy rule-based model is constructed using the following two rules: (20 points)

R1: IF X is slightly A and Y is B THEN Z is not very C

R2: IF X is A and Y is very B THEN Z is C

Where: $A = \left\{ \frac{0.2}{x1} + \frac{0.5}{x2} + \frac{0.9}{x3} + \frac{1}{x4} + \frac{0.7}{x5} + \frac{0.3}{x6} \right\}$

$B = \left\{ \frac{1}{y1} + \frac{1}{y2} + \frac{0.9}{y3} + \frac{0.7}{y4} + \frac{0.4}{y5} + \frac{0.1}{y6} \right\}$

$C = \left\{ \frac{0.1}{z1} + \frac{0.4}{z2} + \frac{0.6}{z3} + \frac{0.8}{z4} + \frac{1}{z5} + \frac{1}{z6} \right\}$

Let the input to the fuzzy model be: ***X is more-or-less A* and Y is very B****, where:

$A^* = \left\{ \frac{1}{x1} + \frac{0.8}{x2} + \frac{0.6}{x3} + \frac{0.45}{x4} + \frac{0.25}{x5} + \frac{0.15}{x6} \right\}$

$B^* = \left\{ \frac{0.1}{y1} + \frac{0.3}{y2} + \frac{0.5}{y3} + \frac{1}{y4} + \frac{0.7}{y5} + \frac{0.2}{y6} \right\}$

Assuming a fuzzy Mamdani model, get the following:

a) The firing degree of each rule.

b) The output possibility distribution of the model.

c) If the output space C={100,200,300,400,500,600}, then obtain the final crisp output.

Q3: Consider a linguistic variable x has the following three linguistic values defined using the following family of fuzzy sets:

$$A_i(x: c_i, w_i) = 1 - \min\left(1, \frac{|x-c_i|}{w_i}\right), \quad x \in [0,8], \quad i = 1,2,3$$

Where : "Low" = $A_1(x: c_1, w_1) = A_1(x: 0, 5)$

"Medium" = $A_2(x: c_2, w_2) = A_2(x: 4, 2)$

"High" = $A_3(x: c_3, w_3) = A_3(x: 8, 6)$

a) Sketch the above linguistic values.

b) Write down closed formulas for the complement of each linguistic value.

c) If x=5.2, get all possibility measures $POS(A_i/x)$, i=1, 2, 3.

d) Get all the cardinalities $|A_i|$, i=1, 2, 3.

e) Get all $COA(A_i)$, i=1, 2, 3. (15 points)

Q4: a) Two triangular fuzzy numbers are defined as follows: A=(2, 3, 9) , B=(3, 5, 6).

Compute the relative rank of both numbers using the methods of:

i) Adamo ii) Yager. iii) Centroid (15 points)

b) In what is called: "Similarity based fuzzy model", the firing degree computed for each rule equals the similarity degree between the input and the fuzzy concept stated in the condition part of the rule. The similarity degree between any two fuzzy sets A, B is computed as: S(A,B) = 1/(1+D(A,B)), where D(A,B) is the distance between the two fuzzy sets. The firing degree of the fuzzy rule is simply computed as the intersection of the all similarity degrees of the inputs with that rule. The output of the fuzzy rule is a crisp value which is obtained as the product of the firing degree of the rule with the prototypical element of the output fuzzy set. Based on the above discussion consider the following fuzzy rule:

IF x_1 is A_1 and x_2 is A_2 THEN y is C

Where: A_1= (2,3,5) , A_2=(1,2,6) , C=(3,5,6) , Get the output of the rule for the following two case:

i- x_1= 3.6 and x_2=1.9

ii- x_1= (1,2.5,4) and x_2=(3,4,7) (20 points)

(Note: Use the vertex method to compute the distance)

Cairo University
Institute of Statistical Studies and Research
Computer and Information Sciences Department

Advanced Topics in CS (CS605)
Selected Topics in IS (IS610)
June 2008, 3 Hours, 3 Pages

Please answer all questions.

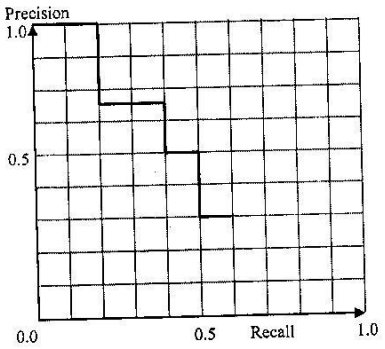
Note: computation of actual final results is not necessary, e.g. you may answer 4+5 or 9.

1. Assume a document retrieval system produced the following interpolated precision-recall curve on a particular query (based on 20 hits). You know that there are 10 relevant documents in the collection.

   a) What is the precision after the system has retrieved three relevant documents?

   b) Going down the hit list, you discover that you've retrieved n documents, and all of them are relevant. What's the maximum possible value of n?

   c) .What are the positions of the relevant documents in first run?



2. An information retrieval system returns the following ranked list for a particular query:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
R	N	R	N	N	N	N	R	R	N	N	N	N	N	R	N	N	N	N	R

   Where Rs represent relevant documents; Ns represent irrelevant documents. From the known relevance judgments, there are 8 relevant documents out of 1000. Any relevant documents not found in the top 20 are assumed to have zero precision. Calculate the following measures:

   a) Mean Average Precision (MAP).

   b) The largest and smallest possible error in MAP caused by the above assumption.

   c) R-precision.

   d) Precision at 10.

   e) Interpolated precision at recall 30%.

   f) F_1 assuming unranked retrieval set.

3. Are the following statements right or false? Justify your answer.

   a) Stemming increases retrieval precision.

   b) Stemming and lemmatization are nearly the same.

   c) Stop lists contains all most frequent terms.

4. Assume the following fragments comprise your document collection:

   Doc 1: Interest rates and rising home costs

   Doc 2: Kids do not have an interest in banking

   Doc 3: Feds' interest in raising interest rates rising

   Doc 4: Kids home costs

   a) Construct the term-document incidence matrix for the above documents that can be used in Boolean retrieval. Assume the following are stop words: *an, and, do, in, not*. You can fill the following table; fill only non-zero cells.

Term	Doc 1	Doc 2	Doc 3	Doc 4
banking				
costs				
feds				
have				
home				
interest				
kids				
raising				
rates				
rising				

   b) Show the positioned postings lists for terms: *kids*, *interest*, *rates* and *rising*. Show the document frequency in your answer. You can use document number such as 1, 2, etc.

   c) Using the incidence matrix, what documents would be returned in response to the following queries? Show the steps of your answer.

      i. `interest AND NOT rates`

      ii. `(interest AND rates) AND NOT (rising OR kids)`

   d) Using postings lists, propose an order for optimal processing of the second query above.

   e) Using postings lists, how the phrase query `"interest rates"` is processed?

   f) Can the time complexity be $O(x+y)$ for merging two postings lists to process the query `(interest AND NOT rates)`? Justify your answer.

   g) What is *tf-idf*? Consider the 3 terms `costs`, `interest`, `rates` only, construct the vector space term-document matrix for the above documents using *tf-idf* term weighting. Normalize your vectors using *Euclidean length*.

   h) Simulate the retrieval of documents in response to the query `interest rats`. Indicate the order in which documents will be retrieved, and the cosine similarity score between the query and each retrieved document.

   i) Consider the query in part (h), assume that only document 1 annotated as relevant. Give the Rocchio-modified query (using $\alpha = 1$, $\beta = 0.75$, and $\gamma = 0.15$.).

5. Give a brief description (or draw a figure) of the architecture of an IR system for text retrieval, supposing Boolean, wildcard and phrase queries are allowed and the documents are ranked.

6. Consider pseudo-relevance feedback and query expansion methods:
   a) Briefly explain the difference between them.
   b) What is effect of each method on precision and recall?

7. Consider wildcard query processing:
   a) What are the permuterm vocabularies the index may include for the word *welcome*?
   b) What Boolean query you would build to search for `wel*o*me` using permuterm index?
   c) What Boolean query you would build to search for the same word using 3-gram index?

Good Luck

Dr. Waleed Arafa