FUZZY SETS AND EXPERT SYSTEMS PROGRAMMING ASSIGNMENT 1
Instructor: Dr. Piero P. Bonissone **Due: October 7, 1999**

Part 1: Generative Grammars

Consider the generative grammar G1 = $\{NT, TE, S, P\}$, where:

NT are the non-terminals {A, B, C, D, E, H, I, N, R, T, U, V}

TE are the terminals {*not, and, or, extremely, very, more_or_less, indeed, hardly, younger_than, older_than, in_between, from_to, young, middle_aged, old, any_age, no_age, unknown_age, (,)*}

P is the set of productions :

$$
\begin{aligned}
S &\rightarrow A|ACA|U \\
A &\rightarrow B|R(A)(A)|IB \\
B &\rightarrow E|NE|DE \\
E &\rightarrow T|HT|VT \\
C &\rightarrow and|or \\
R &\rightarrow in\_between|from\_to \\
N &\rightarrow not \\
D &\rightarrow younger\_than|older\_than \\
I &\rightarrow indeed|hardly \\
V &\rightarrow extremely \\
H &\rightarrow very|more\_or\_less \\
T &\rightarrow young|middle\_aged|old \\
U &\rightarrow any\_age|no\_age|unknown\_age
\end{aligned}
$$

1. (a) How many *different* sentences are there in the language $L(G1)$ generated by the grammar G1 ?

   (b) Use a random (or manual) procedure to fire a subset of rules in P and generate 60 sentences (and their corresponding production tree) out of the above language. This is a subset of $L(G1)$. Call it **S1**.

   (c) Store the subset **S1** of sentences and then evaluate the semantics of each sentence. To do this, you must have implemented a semantic interpreter. (Instructions for this are given in Part 2). If you wish, you may augment the above grammar with additional parentheses to represent the structural information derived from the production tree of each sentence.

2. Select 15 meaningful sentences out of S1 (call this S11) and show their meaning, i.e., their membership distributions, in a tabular or graphical form.

3. Out of the 60 sentence subset S1, did you find any two (or more) sentences with the *same* meaning (i.e., the same membership distribution), thus forming an equivalence class? If you did not find them in your 40 sentence sample, do you believe it is possible to find non-trivial equivalence classes (i.e., classes with more than one element in it) in the language $L(G1)$?

4. Out of the 60 sentence subset S1, did you find any *meaningless* sentence (i.e., any sentence with an empty meaning), which would therefore be equivalent to *no_age*? If you did not find it in your 40 sentence sample, do you believe it is possible to have such a case in the language $L(G1)$ (excluding of course the trivial case of *no_age*)?

5. Define a grammar G2, in the same manner as G1 was defined, such that $|L(G2)| \ll |L(G1)|$ and such that most sentences in $L(G1)$ could be approximated (meaningwise) by the sentences in $L(G2)$. If you look at the slide on *granularity* you might get some ideas about how to cover all the values of the universe of discourse with fuzzy sets of different levels of granularity. (You do not have to implement $G2$, unless you want to answer the associated optional question later.)

6. Make the following three changes, and then redo Q.1-c and Q.2 above, using the same sets S1 and S11 that you used before (i.e. evaluate the meanings of sentences in S1 and show the membership functions of the sentences in S11).

   - Decrease the interval of the universe of discourse (Part 2-Step 1) to the interval [0,90].
   - Instead of using the S and Π functions, use the linear functions and the four parameter representation (Part 2-Step 2.2) to define the primary terms *young, middle_aged, old.* Let these terms be defined by (0,28,0,12), (40,58,12,7), (65,90,7,0) respectively.
   - Increase the number of sampling points N by 50% (Part 2-Step 3).

7. Defuzzify the fuzzy sets defined by sentences in S11. Use the Center-of-Gravity (first moment) method on the vector representation. For each, we will get a single number which represents the fuzzy set. This number could be very atypical of that set — characterize those properties of a membership distribution which when present, lead to a misleading and unreasonable defuzzification result. Can you suggest a defuzzification method which will not have these problems?

   Optional Questions :

8. Modify Q.1-b as follows: instead of generating the sentences from the grammar, develop a *parser* based on the same grammar. The parser will either accept or reject a sentence input to it. If the sentence is accepted, the parser will provide the semantic interpreter the relevant structural information (the production or derivation tree) to facilitate the sentence's interpretation.

9. Modify Q.6-c as follows: instead of using an increased number of sampling points, avoid sampling completely. Use the linear functions and the four parameter representation (a, b, α, β) as the underlying representation of the membership distributions. Use the alternative set

of definitions for hedges, relations, and other terminals in the grammar. Extend your data structures to handle subnormal fuzzy sets (caused by the *and*, *not*) and non-convex fuzzy sets (caused by *or*). Answer Q.6 again, after these changes.

10. Implement the linguistic approximation of the 15 sentence set S11 (defined in Q.2) using sentences in $L(G2)$ to provide the description of each of the original sentences. In order to determine when two membership functions are "close", we need a metric to measure the distance between them. You can use 2 parameters (area and first moment) for prescreening and the L_2 or Euclidean norm for final label selection. $(L_2(f,g) = \sigma_i(f(u_i) - g(u_i))^2)$.

11. Compute the analytic formula for the defuzzification of a fuzzy set represented (without sampling) as (a) $S(u;\alpha,\beta,\gamma)$, (b) $\Pi(u;\beta,\gamma)$, (c) (a,b,α,β), using each of the methods: (i) Center-of-Gravity, (ii) Mean-of-Maximum. See Appendix for definitions.

12. Given the following fuzzy numbers, defined in terms of their four parameter representation (a,b,α,β):

    - A=(1.0,1.5,0.2,0.1)
    - B=(0.5,0.6,0.1,0.1)
    - C=(3,3,0,0)
    - D=(4.5,5,0,0.2,0.2)

    implement the fuzzy artihmetics formulae (from lecture 3 handout) and obtain the results of the following expressions:

    (a) $(3A + B)/D$
    (b) $1/(A^2 + D)$
    (c) $(A * B) + [C * (1 - B)] + (B/D)$
    (d) $(A - B)/C$
    (e) A^B
    (f) B^C

Part 2 : Semantic Interpretation

The Semantic Interpreter must be able to compute the meanings of syntactically correct sentences. For doing this, you need to make some assumptions and implementation choices, such as the universe of discourse, the representational form of membership functions, how they will be stored, and what the transformation rules are. Follow these steps :

Step 1 Choose the interval for the universe of discourse, i.e. [w,z]. By changing the values of w and z you should be able to map various universes of discourse to your internal representation. For this problem, a reasonable choice is [0, 100].

Step 2 Define the membership distribution of each primary term, i.e. of each terminal element $\in$ TE which can be generated by non-terminal T. This can be done is several ways :

(a) For Part 1 – Q.1-b, use the S and Π functions and parameters α, β, γ. For example,

- *young* = 1 - S(u;0,25,50),
- *middle_aged* = Π(u;18,42),
- *old* = S(u;50,65,80).

(b) For Part 1 – Q.6, use the linear functions and the 4-parameter representation (a, b, α, β).

- *young* = (0,28,0,12),
- *middle_aged* = (40,58,12,7),
- *old* = (65,90,7,0).

The appendix provides the definition of both sets of functions. In either case, assign either the suggested or your own values to define the three primary terms. We also need to define the membership distribution of the terms that can be generated by non-terminal U, i.e., *any_age*, *no_age*, and *unknown_age*. The first two are the *universal* and the *empty* set, respectively. The third one correspond to the least informative age value, i.e, the set that maximizes the measure of fuzziness (entropy or equivalent). Thus, the membership distributions of these three sets should be:

- $\mu_{any\_age}(u_i) = 1$ for $i = 1, \ldots, n$
- $\mu_{no\_age}(u_i) = 0$ for $i = 1, \ldots, n$
- $\mu_{unknown\_age}(u_i) = 0.5$ for $i = 1, \ldots, n$

Step 3 All membership functions could be stored as vectors, containing the values $\mu()$ over a discretized universe of discourse. Determine the number of points in which the interval [w,z] will be uniformly sampled. Call this number N. (Common sense questions: Which criteria should N satisfy? When would non-uniform sampling be a better idea?)

Step 4 Each *primary term* (derivable from non-terminal T) and each extreme term (derivable from non terminal U) is now a N-dimensional vector (unless you are implementing the optional Part 1–Q.9, in which case see Step 5 later in this part). Let X be the N-dimensional vector corresponding to:

$$\langle \mu_X(u_1), \mu_X(u_2), ..., \mu_X(u_n) \rangle$$

All the vector operations used in the definitions are to be computed element-wise, so that $f(X)$ is equivalent to applying $f()$ to each element $\mu_X(u_i)$ for each i, e.g.:
$X^2 = \langle ...(\mu_X(u_i))^2 ... \rangle$
$1 - X = \langle ...1 - \mu_X(u_i) ... \rangle$
Max $\{X, Y\} = \langle ... \max(\mu_X(u_i), \mu_Y(u_i)) ... \rangle$

Definition of Semantics

X *and* Y	min {X, Y}
X *or* Y	max {X, Y}
not X	1 - X
very X	CONTR(X) = X^2

extremely X	CONTR(CONTR(X)) = X^4
more_or_less X	DIL(X) = $X^{0.5}$
indeed X	INT(X)
hardly X	FUZ(X)
younger_than X	SM(X)
older_than X	GR(X)
in_between (X)(Y)	NORM(GR(X) *and* SM(Y))
	= NORM((*not* SMEQ(X)) *and* (*not* GREQ(Y)))
from_to (X)(Y)	GREQ(X) *and* SMEQ(Y)

where:

NORM(X)	= X / $\max_i \mu_X(u_i)$
INT(X)	= $2X^2$ when $X < 0.5$
	= $1 - 2(1-X)^2$ when $X > 0.5$
FUZ(X)	= $0.5 - 2(0.5 - X)^2$ when $X < 0.5$
	= $0.5 + 2(X - 0.5)^2$ when $X > 0.5$
SM(X)	= *not* GREQ(X)
GR(X)	= *not* SMEQ(X)
GREQ(X)	= X for $u_i < u^*$
	= 1 for $u_i > u^*$
SMEQ(X)	= 1 for $u_i < u^*$
	= X for $u_i > u^*$

Here, $u^* = \min\{u | \mu_X(u) = 1\}$ (the leftmost value of U with membership 1)

NOTE: if necessary, normalize $\mu_X(u)$ before applying SMEQ(X) or GREQ(X).

(OPTIONAL) STEP 5: (NEEDED ONLY FOR PART 1 – Q.9)

Each *primary term* is now a four parameter list. Let X = (a, b, α, β) and Y = (c, d, γ, δ). Let the universe of discourse be the interval [w,z].

For this optional step, disregard the terms derivable from non-terminal U, because *no_age* or *unknown_age* are not normal fuzzy numbers. Therefore, they cannot be represented by the four parameter representation.

Definitions

$$
\begin{array}{lll}
X\ and\ Y & = (p, q, L_a, R_a), & \text{where} \\
p & & = \max(a, c) \\
q & & = \min(b, d) \\
L_a & & = (c + \alpha) - min(a, c), \quad (a - \alpha) > (c - \gamma) \\
 & & = (a + \gamma) - min(a, c), \quad (a - \alpha) < (c - \gamma) \\
 & & = \alpha, \quad (a - \alpha) = (c - \gamma), a > c \\
 & & = \gamma, \quad (a - \alpha) = (c - \gamma), a < c \\
 & & = \alpha = \gamma, \quad (a - \alpha) = (c - \gamma), a = c \\
R_a & & = (d + \delta) - min(b, d), \quad (b + \beta) > (d - \delta) \\
 & & = (b + \beta) - min(b, d), \quad (b + \beta) < (d - \delta) \\
 & & = \delta, \quad (b + \beta) = (d - \delta), b > d \\
 & & = \beta, \quad (b + \beta) = (d - \delta), b < d \\
 & & = \beta = \delta, \quad (b + \beta) = (d - \delta), b = d
\end{array}
$$

Note: If max(a,c) $\leq$ min(b,d), the result is still a *normal* set. Otherwise, one should normalize the result, by finding the intersection of the two slopes, i.e.:

X *and* Y = NORM$(p, q, L_a, R_a) = (x, x, L_n, R_n)$, where

$x = (qR_a + pL_a)/(L_a + R_a)$

$L_n = x - (p - L_a)$

$R_n = (q + R_a) - x$

$$
\begin{array}{llll}
X \text{ } or \text{ } Y & = (r, s, L_o, R_o), & & \\
 & r & = min(a, c) & \\
 & s & = max(b, d) & \\
 & L_o & = (a+\gamma) - \max(a, c), & (a-\alpha) > (c-\gamma) \\
 & & = (c+\alpha) - \max(a, c), & (a-\alpha) < (c-\gamma) \\
 & & = \gamma, & (a-\alpha) = (c-\gamma), a > c \\
 & & = \alpha, & (a-\alpha) = (c-\gamma), a < c \\
 & & = \alpha = \gamma, & (a-\alpha) = (c-\gamma), a = c \\
 & R_o & = (b+\beta) - \max(b, d), & (b+\beta) > (d-\delta) \\
 & & = (d+\delta) - \max(b, d), & (b+\beta) < (d-\delta) \\
 & & = \delta, & (b+\beta) = (d-\delta), b > d \\
 & & = \beta, & (b+\beta) = (d-\delta), b < d \\
 & & = \beta = \delta, & (b+\beta) = (d-\delta), b = d
\end{array}
$$

Note: If max(a,c) $\leq$ min(b,d) the result is still a *convex* set. Otherwise, just maintain both fuzzy sets as elements of a data structure called UNION-LIST, i.e., X *or* Y = UNION-LIST(X,Y). Any subsequent operation on the result of X *or* Y can be distributed to each element in UNION-LIST.

The remaining definitions are :

not X	= UNION-LIST(SM(X), GR(X))	
very X	= CONTR(X) = $(a+\alpha/5, b-\beta/5, \alpha, \beta)$	if $(a+\alpha/5) \leq (b-\beta/5)$
	= NORM(CONTR(X))	otherwise
extremely X	= CONTR(CONTR(X)) = $(a+\alpha/2.5, b-\beta/2.5, \alpha, \beta)$	if $(a+\alpha/2.5) \leq (b-\beta/2.5)$
	= NORM(CONTR(CONTR(X)))	otherwise
more_or_less X	= DIL(X) = $(a-\alpha/5, b+\beta/5, \alpha, \beta)$	
indeed X	= INT(X) = $(a-\alpha/4, b+\beta/4, \alpha/2, \beta/2)$	
hardly X	= FUZ(X) = $(a+\alpha/4, b-\beta/4, \alpha/2, \beta/2)$	if $(a+\alpha/4) < (b-\beta/4)$
	= $(a+\alpha(b-a)/(\alpha+\beta), (b-\beta(b-a)/(\alpha+\beta), \alpha/2, \beta/2)$	otherwise
younger_than X	= SM(X) = $(w, a-\alpha, 0, \alpha)$	
older_than X	= GR(X) = $(b+\beta, z, \beta, 0)$	
from_to (X)(Y)	= (a, d, α, δ)	
in_between (X)(Y)	= *from_to* (GR(X))(SM(Y)) = $(b+\beta, c-\gamma, \beta, \gamma)$	

NOTE: if necessary, normalize X before applying SM(X) or GR(X).

APPENDIX

A computationally efficient way to characterize a fuzzy set is to use a parametric representation of its membership function. Here we give two possible parametric representations. Remember that regardless of how we define it, we still have a choice of implementing the membership function (in software) as a sampled vector or as a function call. The sampled form is less accurate but more efficient.

S and Π functions

A fuzzy set can be represented by $S(u;\alpha,\beta,\gamma)$ or $\Pi(u;\beta,\gamma)$, where the S and Π functions are defined as:

$$S(u;\alpha,\beta,\gamma)=\begin{cases}0 & \text{if } u\le\alpha\\ 2((u-\alpha)/(\gamma-\alpha))^2 & \text{if } \alpha\le u\le\beta\\ 1-2((u-\gamma)/(\gamma-\alpha))^2 & \text{if } \beta\le u\le\gamma\\ 1 & \text{if } u\ge\gamma\end{cases}$$

In the definition of the S function, parameter β is defined as $\beta=(\alpha+\gamma)/2$ and represents the cross-over point, i.e. $S(\beta)=0.5$.

$$\Pi(u;\beta,\gamma)=\begin{cases}S(u;\gamma-\beta,\gamma-(\beta/2),\gamma) & \text{if } u\le\gamma\\ 1-S(u;\gamma,\gamma+(\beta/2),\gamma+\beta) & \text{if } u\ge\gamma\end{cases}$$

In the definition of the Π function, parameter β indicates the length of the bandwidth, while parameter γ indicates the point at which Π is normal, i.e. $\Pi(\gamma)=1$.

Linear function representations

Another alternative is the four parameter representation (a,b,α,β) for a fuzzy set. The first two parameters indicate the interval of the universe of discourse in which the membership value is 1.0; the third and fourth parameters indicate the left and right *width* of the distribution. Line segments are used to define the slopes. Therefore, the membership function of such a fuzzy set A is defined as:

$$\mu_A(u)=\begin{cases}0 & \text{if } u<(a-\alpha)\\ (1/\alpha)(u-a+\alpha) & \text{if } u\in[(a-\alpha),a]\\ 1 & \text{if } u\in[a,b]\\ (1/\beta)(b+\beta-u) & \text{if } u\in[b,(b+\beta)]\\ 0 & \text{if } u>(b+\beta)\end{cases}$$

Recall that we have already seen this representation for fuzzy numbers, and the formulae for manipulating them for fuzzy arithmetic operations.

Defuzzification

- Center-of-Gravity Method

  The center of gravity (or centroid of the distribution) is obtained by computing the first moment of $\mu_X(z)$:

  $$\text{DEFUZ}_{COG}(X) = \frac{\int_U z\mu_X(z)dz}{\int_U \mu_X(z)dz}$$

  For the discrete and finite case, where U=$z_1, z_2, \ldots, z_n$, this expression reduces to:

  $$\text{DEFUZ}_{COG}(X) = \frac{\sum_{i=1,n} z_i\mu_X(z_i)}{\sum_{i=1,n} \mu_X(z_i)}$$

- Mean Of Maxima

  The MOM method obtains the mode of the distribution. i.e the value of z in which the membership distribution $\mu_X(Z)$ achieves its maximum. If the maximum is obtained in multiple points, the output is the average of such set of points.

  Using the Mean-of-Maximum method, we first define $V \subset U$ as

  $$V = \{z|\mu_X(z) \geq \mu_X(t), \quad z, t \in U\}$$

  That gives the defuzzification as

  $$\text{DEFUZ}_{MOM}(X) = \frac{\int_V z\mu_X(z)dz}{\int_V \mu_X(z)dz} = \frac{\int_V zdz}{\int_V dz}$$

  For the discrete and finite case, where U=$z_1, z_2, \ldots, z_n$, we first define the set of points $V \subset U$ as

  $$V = \{z_i|\mu_X(z_i) = \bigvee_{j=1,n} \mu_X(z_j)\}$$

  and then we define output of the MOM defuzzification as the average of the points in V:

  $$\text{DEFUZ}_{MOM}(X) = \frac{\sum_{z_i \in V} z_i\mu_X(z_i)}{\sum_{z_i \in V} \mu_X(z_i)} = \frac{\sum_{z_i \in V} z_i}{\mid V \mid}$$