Lecture 16: Uncertainty - 6

Victor Lesser

CMPSCI 683
Fall 2004

Today's lecture

- Alternative Models of Dealing with Uncertainty Information/Evidence
  - Dempster-Shaffer Theory of Evidence
  - Fuzzy logic
  - Logical ways of dealing with uncertainty

Dempster-Shafer Theory

- **Designed to Deal with the distinction between uncertainty and ignorance**
  - Rather than computing the probability of a proposition it computes the probability that evidence **supports** the proposition
- **Applicability of D-S**
  - assume lack of sufficient data to accurately estimate the prior and conditional probabilities to use Bayes rule
  - incomplete model ⇒ rather than estimating probabilities it uses belief intervals to estimate how close the evidence is to determining the truth of a hypothesis

Dempster-Shafer Theory

Allows representation of *ignorance* about support provided by evidence

- allows reasoning system to be skeptical

For example, suppose we are informed that one of three terrorist groups, A, B or C has planted a bomb in a building.

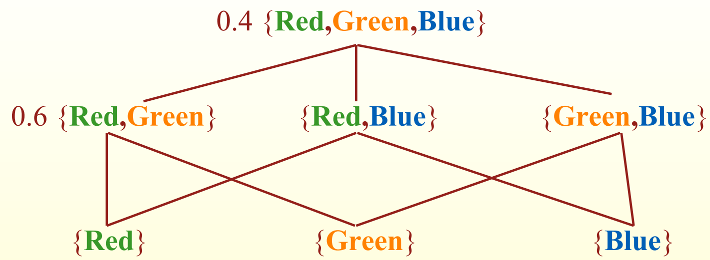
We may have some evidence the group C is guilty, P (C) =0.8

We would not want to say the probability of the other two groups being guilty is .1

In traditional theory, forced to regard belief and disbelief as functional opposites p(a) + p(not a) = 1 and to distribute an equal amount of the remaining probability to each group

D-S allows you to leave relative beliefs unspecified

Belief Subsets



Suppose that the evidence supports {red,green} to the degree .6. The remaining support will be assigned to {red,green,blue} while a Bayesian model assumes that the remaining support is assigned to the negation of the hypothesis (or its complement) {blue}.

Dempster-Shafer theory

- **Given a population F =(blue,red, green) of mutually exclusive elements, exactly one of which (f) is true, a basic probability assignment (m) assigns a number in [0,1] to every subset of F such that the sum of the numbers is 1.**
  - Mass as a representation of evidence support
- **There are $2^{|F|}$ propositions, corresponding to "the true value of f is in subset A".**
  - (blue),(red),(green),(blue, red), (blue,green),(red,green),(red,blue,green),(empty set)
- **A belief in a subset entails belief in subsets containing that subset.**
  - Belief in (red) entails Belief in (red,green),(red,blue),(red,blue,green)

Interpretation of m(X)

- **Random switch model**: Think of the evidence as a switch that oscillates randomly between two or more positions. The function *m* represents the fraction of the time spent by the switch in each position.
- **Voting model**: *m* represents the proportion of votes cast for each of several results of the evidence (possibly including that the evidence is inconclusive).
- **Envelope model**: Think of the evidence as a sealed envelope, and *m* as a probability distribution on what the contents may be.

Belief and Plausibility

- **Belief (or possibility) is the probability that B is provable (supported) by the evidence.**
  - $Bel(A) = \Sigma_{\{B \text{ in } A\}} m(B)$ (Support committed to A)
  - *Bel ((red,blue)) = m((red))+m((blue))+m((red,blue))*
- **Plausibility is the probability that B is compatible with the available evidence (cannot be disproved).**
  - Upper belief limit on the proposition A
- **$Pl(A) = \Sigma_{\{B \cap A \neq \{\}\}} m(B)$ Support that can move into A**
  - *Pl((red,blue)) = m((red))+m((blue))+m((red,blue))+m((red,green))+ m((red,blue,green))+m((blue,green))*
- **$Pl(A) = 1\text{-}Bel(\neg A)$**

Confidence Interval

- Belief Interval [Bel(A),Pl(A)], confidence in A
  - **Interval width is good aid in deciding when you need more evidence**
- [0,1] no belief in support of proposition
  - **total ignorance**
- [0,0] belief the proposition is false
- [1,1] belief the proposition is true
- [.3,1] partial belief in the proposition is true
- [0,.8] partial disbelief in the proposition is true
- [.2,.7] belief from evidence both for and against propostion

DS's Rule of Combination

- **Given two basic probability assignment functions m_1 and m_2 how to combine them**
  - **Two different sources of evidence**
  - **$P(s_i,s_j|d) = P(s_i|d)\ P(s_j|d)$; assume conditional independence**
- **Bel (C) = Sum ($m_1(A_i)\ m_2(B_j)$) where *A_i intersect B_j = C***
- **Normalized by amount of non-zero mass left after intersection Sum ($m_1(A_i)\ m_2(B_j)$) where A_i intersect B_j not empty**

Example of Rule Combination

- Suppose that $m_1(D)=.8$ and $m_2(D)=.9$

	{D} .9	{¬D} 0	{D,¬D} .1
{D} .8	{D} .72	{} 0	{D} .08
{¬D} 0	{} 0	{¬D} 0	{¬D} 0
{D,¬D} .2	{D} .18	{¬D} 0	{D,¬D} .02

- $m_{12}(D) = .72+.18+.08=.98$ $m_{12}(\neg D)=0$ $m_{12}(D,\neg D)=.02$
- Using intervals: [.8,1] and [.9,1] = [.98,1]

DS's Rule of Combination cont.

- Suppose that $m_1(D)=.8$ and $m_2(\neg D)=.9$

	{D} 0	{¬D} .9	{D,¬D} .1
{D} .8	{D} 0	{} .72	{D} .08
{¬D} 0	{} 0	{¬D} 0	{¬D} 0
{D,¬D} .2	{D} 0	{¬D} .18	{D,¬D} .02

- Need to normalize (.18+.08+.02) by .72:
  $m_{12}(D) = .29$ $m_{12}(\neg D)=.64$
  $m_{12}(D,\neg D)=.07$
- Using intervals: [.8,1] and [0,.1] = [.29,.36]

Dempster-Shafer Example

Let *Θ* be:

All : allergy
Flu : flu
Cold : cold
Pneu : pneumonia

When we begin, with no information *m* is:

{*Θ* } (1.0)

suppose m_1 corresponds to our belief after observing fever:

{*Flu, Cold, Pneu* } (0.6)
{*Θ* } (0.4)

Suppose m_2 corresponds to our belief after observing a runny nose:

{*All, Flu, Cold*} (0.8)
Θ (0.2)

Dempster-Shafer Example (cont'd)

Then we can combine m_1 and m_2:

		{A,F,C }	(0.8)	Θ	(0.2)
{F,C,P }	(0.6)	{F,C }	(0.48)	{F,C,P}	(0.12)
Θ	(0.4)	{A,F,C }	(0.32)	Θ	(0.08)

So we produce a new, combined m_3;

{Flu,Cold }	(0.48)
{All,Flu,Cold }	(0.32)
{Flu,Cold,Pneu }	(0.12)
Θ	(0.08)

Suppose m_4 corresponds to our belief that the problem goes away on trips and thus is associated with an allergy:

{All}	(0.9)
Θ	(0.1)

Dempster-Shafer Example (cont'd)

Applying the numerator of the combination rule yields:

		{A}	(0.9)	Θ	(0.1)
{F,C}	(0.48)	{}	(.432)	{F,C}	(0.048)
{A,F,C}	(0.32)	{A}	(0.288)	{A,F,C}	(0.032)
{F,C,P}	(0.12)	{}	(.108)	{F,C,P}	(0.012)
Θ	(0.08)	{A}	(0.072)	Θ	(0.008)

Normalizing to get rid of the belief of 0.54 associated with {} gives m_5:

{Flu,Cold }	(0.104)=.048/.46
{Allergy,Flu,Cold }	(0.0696)= .032/.46
{Flu,Cold,Pneu }	(0.026)= .012/.46
{Allergy }	(0.782)=(.288+.072)/.46
Θ	(0.017)=.008 /.46

What is the [belief, possibility] of Allergy?

Dempster-Shafer Pros

- Addresses questions about necessity and possibility that Bayesian approach cannot answer.
- Prior probabilities not required, but uniform distribution cannot be used when priors are unknown.
- Useful for reasoning in rule-based systems

Dempster-Shafer Cons

- Source of evidence not independent; can lead to misleading and counter-intuitive results
- The normalization in Dempster's Rule loses some meta-information, and this treatment of conflicting evidence is controversial.
  - **Difficult to develop theory of utility since Bel is not defined precisely with respect to decision making**
- Bayesian approach can also do something akin to confidence interval by examining how much one's belief would change if more evidence acquired
  - **Implicit uncertainty associated with various possible changes**

Conflicting Evidence and Normalization Problems with Dempster-Shafer Theory

Normalization process can produce strange conclusions when conflicting evidence is involved:

Θ = {A,B,C}

m_1 = {A} (0.99), {B} (0.01)

m_2 = {C} (0.99), {B} (0.01)

$m_1 + m_2$ = {B} (1.0)

- **Certain of B even though neither piece of evidence supported it well**
- **No representation of inconsistency and resulting uncertainty/ ignorance**

Fuzzy Set Theory/Logic

- **Method for reasoning with logical expressions describing fuzzy set membership**
  - Knowledge representation based on *degrees of membership* rather than a crisp membership of binary logic
    - Rather than stochastic processes
  - *Degree of truth in proposition rather than degree of belief*
- ***Logic of gradual properties*** **as well as calculus for incomplete information**
  - "precision in reasoning is costly and should not be pursued more than necessary."
- **Application - Fuzzy Control Theory**
  - **expert knowledge coded as fuzzy rules**
    - **if the car's speed is *slow* then the braking force is *light***
  - **computer control of a wide range of devices**
    - **washing machines, elevators, video cameras, etc.**

Representing Vagueness/Fuzziness

- **Fuzziness is a way of defining concepts or categories that admit vagueness and degree**
  - nothing to do with degree of belief in something and need not be related to probabilities
    - we believe (.5) it will rain today
  - example of "was it a rainy day" fuzziness
    - misty all day long but never breaks into a shower
    - rain for a few minutes and then sunny
    - heavy showers all day long

Example of Fuzzy Control Rules

Given two inputs:

E = difference between current temperature and target temperature normalized by the target temperature.

dE = the time derivative of E (dE/dt)

Compute one output:

W = the change in heat (or cooling) source

Fuzzy variables: NB (neg big), NS, ZO, PS, PB

Example rule: if E is ZO and dE is NS then W is PS

if E IS ZO and dE is PB then W is NB

....

Significant Reduction in number of rules needed and handles noisy sensors

Fuzzy Logic

- A different type of approximation related to "vagueness" rather than "uncertainty."
- Measures the "degree of membership" in certain sets or categories such as: age, height, red, several, old, many,...
- Example: "Several" = {2/.3, 3/.5, 4/1, ..., 9/.1}
- Representation A = {u/a(u) | u ∈ U}

Fuzzy Variable

- **Fuzzy variable takes on a fuzzy set as a value**
  - A fuzzy set (class) A in X is characterized by a membership function v/a that assigns each point x in X a real number between 0,1
- **Height Example-- in the tall class**
  - v/a (x) = 1 for any person over 6 feet tall
  - v/a (x) = 0 for any person under 5 feet tall
  - v/a (x) in between for height > 5 and < 6
- ***Piecewise Linear Function for values in between***
  - vector/tall = (0/4, 0/5, 1/6, 1/ 7)
- **Hedge**
  - systematic modification to a characteristic function to represent a linguistic specialization
  - "very tall" v/very tall (x) = (v/tall (x))^2

Fig. 6.38

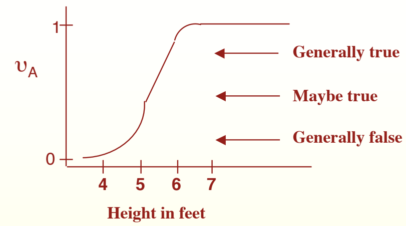


Fig. 6.38:
An example characteristic function for a fuzzy set representing "tall people." The function f_A indicates the degree to which individuals of different heights would be considered to be members of the fuzzy set of tall persons.

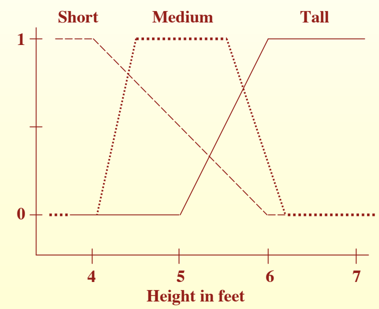


Fig. 6.39:
Three piecewise linear characteristic functions

Possibility Distribution

- Express preferences on possible values of a variable where exact value is not known
  - "concept of medium"
  - Used in reasoning in fuzzy rule sets
- *Consistency of a subset* with respect to a concept "tall people"
  - A={.5/Joe = 5.5, .9/Bob = 5.9, .6/Ray = 5.6}
  - **Possibility Measure ∏(A)**
    - **∏(A) max of v/tall over A ≤ 1**
    - **measure of the degree to which the set "tall people" is possibly in A**
    - .9/Bob = 5.9 implies ∏(A) = .9
  - **Necessity Measure N(A)**
    - **N(A) = 1-∏ (complement of A)**
    - **measure of the degree to which the set "tall people" is necessary in A**
    - Complement of A={.5(1-.5)/Joe = 5.5, .1(1-.9)/Bob = 5.9, .4(1-.6)/Ray = 5.6}
    - **N(A)= 1-.5 = .5**

Fuzzy Set Operators

- **Representation A = {u/a(u) | u ∈ U}**
- **Set operators over same variable:**

A ∪ B = {u/max(a(u), b(u)) | u ∈ U}

A ∩ B = {u/min(a(u), b(u)) | u ∈ U}

¬ A = {u/(1 - a(u)) | u ∈ U}

For example: Young and Rich (Y ∩ R) . . .

cartesian product over two variables:

A x B = {(u,v)/min(a(u), b(v)) | u ∈ U, v ∈ V}

Other set operators are sometimes used, for example:

A ∪ B = {u/(a(u) + b(u) - a(u) · b(u)) | u ∈ U }

A ∩ B = {u/(a(u) · b(u)) | u ∈ U}

etc...

Fuzzy Inference (extension of modus ponens)

Let c, c´, d, d´, be fuzzy sets. Then the "generalized modus ponens" states:

x is c´
<u>**if x is c then y is d;**</u>
y is d´

e.g.:

Visibility is *very* low
<u>If visibility is low then condition is poor</u>
Condition is *very* poor

How c is characterized by c' effect how d is characterized by d'

The fuzzy inference is based on two concepts:

1) fuzzy implication
2) the compositional rule of inference

Fuzzy implication: represented as:

a → b

a and b are fuzzy sets.

Fuzzy Relation

- **A fuzzy relation is represented by a matrix.**
- **A fuzzy relation <u>R</u> associated with the implication**
  - a → b is a fuzzy set of the cartesian product U x V:
  - **R = {(u, v)/m(u,v) | u ∈ U, v ∈ V}**
- **One of the most commonly used fuzzy implications is based on the min op:**
  - $m_R(u,v) = \min(a(u), b(v))$

The Compositional Rule of Inference

If R is a relation from A to B
S is a relation from B to C

The composition of R and S is a relation from A to C denoted R ∘ S

R ∘ S = {(a,c)/max[min(m_R(a,b), m_S(b,c))]}
b

Similarly, can be used as rule of inference when:

- **R is a fuzzy relation from U to V**
- **X is a fuzzy subset of U**
- **Y is a fuzzy subset of V**

Y is induced by X and R by: Y = X ∘ R; modus-ponens

Y = {v/max[min(m_x(u), m_R(u,v))] | u ∈ U}
u

notice that when R = (a → b)

x = a

then y = b

Example of max-min inference

If speed is normal
then braking.force is medium.

speed:
Normal = (0/0, .1/20, .8/40, 1/60, .1/80, 0/100)

braking.force:
Medium = (0/0, .5/1, 1/2, 1/3, .2/4, 0/5)

When characteristic functions are piecewise linear, one way to carry out max-min inference is to define a matrix for the composition. The matrix M for compositional inference is defined so that

$$m_{ij} = \min(a_i, b_j)$$

The matrix for this example is as follows...

Example of max-min inference, p.2

$$m_{ij} = \min(a_i, b_j)$$

The matrix for this example is as follows:

M =	Brake					
Speed	**0**	**1**	**2**	**3**	**4**	**5**
0	min(0,0)	min(0,.5)	min(0,1)	min(0,1)	min(0,.2)	min(0,0)
20	min(.1,0)	min(.1,.5)	min(.1,1)	min(.1,1)	min(.1,.2)	min(.1,0)
40	min(.8,0)	min(.8,.5)	min(.8,1)	min(.8,1)	min(.8,.2)	min(.8,0)
60	min(1,0)	min(1,.5)	min(1,11)	min(1,1)	min(1,.2)	min(1,0)
80	min(.1,0)	min(.1,.5)	min(.1,1)	min(.1,1)	min(.1,.2)	min(.1,0)
100	min(0,0)	min(0,.5)	min(0,1)	min(0,1)	min(0,.2)	min(0,0)

Simplifying terms, we have the following:

0	0	0	0	0	0
0	.1	.1	.1	.1	0
0	.5	.8	.8	.2	0
0	.5	1	1	.2	0
0	.1	.1	.1	.1	0
0	0	0	0	0	0

Example of Max-Min Inference, cont'd

The braking force predicted by max-min inference is the fuzzy vector resulting from multiplying the matrix M by the input speed ***S*: = (0/0, 0/20, .8/40, 0/60, 0/80, 0/100)**

braking.force = B = S ° M

M=	Brake
Speed	.. 2 ..
0	0
20	.1
40	.8
60	1
80	.1
100	0

The general term is given by:

$$b_j = \vee_{1 \le i \le n} [\wedge(s_i\, m_{ij})]$$

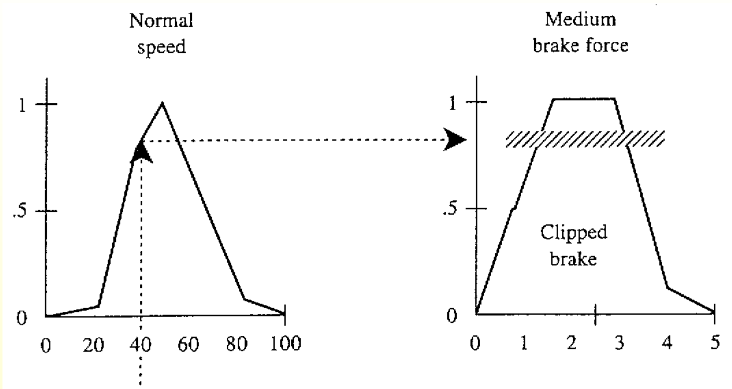
Thus, for the given speed vector, we have:

b_0 = max[min(0,0),min(0,0),min(.8,0), min(0,0), min(0,0), min(0,0)] = 0
b_1 = max[min(0,0),min(0,.1),min(.8,.5), min(0,.5), min(0,.1), min(0,0)] = .5
b_2 = max[min(0,0),min(0, .1),min(.8,.8), min(0,1), min(0,.1), min(0,0)] = .8
b_3 = max[min(0,0),min(0, .1),min(.8,.8), min(0,1), min(0,.1), min(0,0)] = .8
b_4 = max[min(0,0),min(0, .1),min(.8,.2), min(0,.2), min(0,.1), min(0,0)] = .2
b_5 = max[min(0,0),min(0,0),min(.8 ,0), min(0,0), min(0,0), min(0,0)] = 0

B= (0, .5, .8, .8, .2, 0)

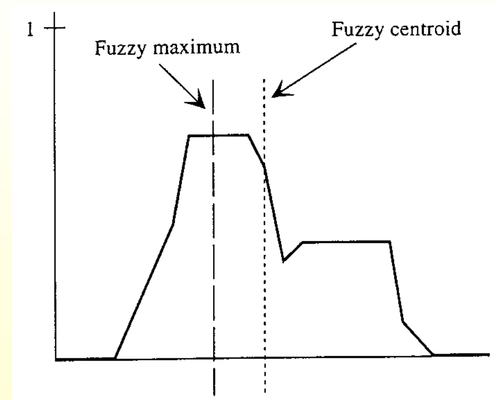
Equivalently, in the vector notation we have the following fuzzy representation for the braking force to be applied: *B* = (0/0, .5/1, **.8/2,** .8/3, .2/4, 0/5) ***Induced fuzzy set***

The effects of max-min inference



Given speed (.8/40)

Two approaches to Defuzzification



How to decide what to do -- what breaking force to apply?

Fuzzy Control

Given two inputs:

E = difference between current temperature and target temperature normalized by the target temperature.

dE = the time derivative of E (dE/dt)

Compute one output:

W = the change in heat (or cooling) source

Fuzzy variables: NB (neg big), NS, ZO, PS, PB

Example rule: if E is ZO and dE is NS then W is PS

if E IS ZO and dE is PB then W is NB

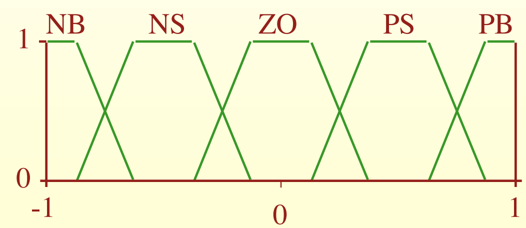
……….

Control rules determine the output

Table Representing All 9 Rules

E \ dE	NB	NS	ZO	PS	PB
NB			PB [6]		
NS			PS [7]		
ZO	PB [1]	PS [2]	ZO [3]	NS [4]	NB [5]
PS			NS [8]		
PB			NB [9]		

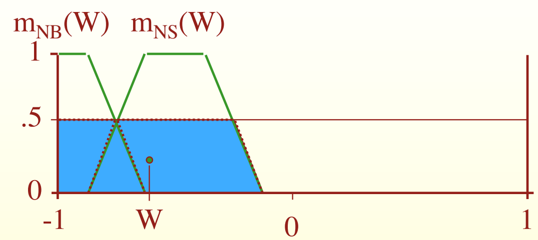
Definition of Variables associated with rules



Computing the Output Value

- Suppose that E=.75 and dE=0
- Hence E is PB with degree .5
  E is PS with degree .5
  dE is ZO with degree 1
- Two rules are applicable:
  $\alpha_8 = m_{PB}(E) \wedge m_{ZO}(dE) = .5 \wedge 1 = .5$
  $\alpha_9 = m_{PS}(E) \wedge m_{ZO}(dE) = .5 \wedge 1 = .5$
- $m_8(W) = \alpha_8 \wedge m_{NS}(W)$
  $m_9(W) = \alpha_9 \wedge m_{NB}(W)$

Computing the Output Value, *cont'd*



- Defuzzification: translating a fuzzy category to a precise output.
- Defuzzification using the center of gravity.
- Why is fuzzy logic so successful?

Why is fuzzy logic so successful?

- Easily and succinctly represent expert system rules that involve continuous variables
- Model environment variables in terms of piece-wise linear characteristic functions

Next Lecture

- More on Logical Reasoning about Uncertainty
- Learning