Uncertainty Management

1. Fuzzy Logic

2. Dempster/Shafer's Theory of evidence

3. Interval Approaches

Fuzzy Set

Let X denote a space of objects.

definition: A *fuzzy set A in X* is a set of ordered pairs

$$A = \{(x, \chi_A(a))\}, x\varepsilon X$$

where $\chi_A(x)$ is termed the *grade of membership of x in A*, which is usually represented by values in [0,1]. That is:

$$\chi_A : X \longrightarrow [0, 1]$$

The grade of membership introduces an ordering in the set A relative to X.

Examples: 1) A="Balanced AHOC-Hand"

Let X be the set of all possible AHOC hands, let $x\varepsilon X$ and let s(x), d(x), h(x) and c(x) denote the number of spades, hearts, diamonds and clubs in x.

$$\chi_A(x) = max(0, 1 - \frac{|s(x) - 3| + |h(x) - 3| + |d(x) - 3| + |c(x) - 3|}{10})$$

remark: The definition of fuzzy set is subjective and usually requires expertise about the use of the concept approximated by the fuzzy set.

2) B="Trump-rich AHOC-Hand"

$$\chi_B(x) = \ If \ sp(x) > 9 \ then \ 1 \ else \ max(0, 1 - \frac{10 - sp(x)}{8})$$

Special problems:
1) Is x trump-rich **and** balanced?
2) Is x trump-rich **or** balanced?
3) **Is there** a hand that is trump-rich and balanced?
4) Are **all** trump-rich hands unbalanced?

Why is Fuzzyness Important in AI-Research?

- Many concepts don't have precise criteria for deciding if an object belongs to this concept. This inexactness can be caused by:
    1) The *generality* of the concept; frequently, the same concept is used in various contexts.
    2) The *ambiguity* of the concept; the same concept has more than one meaning.
    3) The *vagueness* of the concept: there are no precise boundaries.

    For example: How can we describe the meaning of: *a long street,a real numbers much greater than 1*, *balanced hand*, *fever* in computers?
- Furthermore, not only the concepts itself are frequently fuzzy but also our natural language uses a large number of fuzzy quantifiers: see for example; *most, usually, rarely, frequently, much, occasionally.*
    (1) Most Americans drive a car
    (2) Most car drivers are aggressive.
    $\vdash$
    (3) Usually Americans are aggressive.

    (1) Most Swedes have blond hair.
    (2) Males rarely have blond hair.
    $\vdash$
    (3) Male Swedes have occasionally blond hair.

    The main objective of fuzzy Logic and the other approaches to uncertainty management is to provide uncertainty models and inference rules based on these models suitable to automate the above inferences.
- The terminology used in the application area frequently involves fuzzy concepts and the expert frequently uses fuzzy quatifiers, when he describes the domain-specific knowledge. Therefore, its adequate representation in computers and the automation of inference based on this fuzzyness is one of the most critical problems when developing a knowledge based system.
- Frequently, it is much easier to approximate fuzzy concepts using funtions instead of explicitly enumerating the degree of membership of each element of the fuzzy set.

Estimation of Membership Functions

pages 9,10 and 28

Example Fuzzy Reasoning

Let X={$John, Fred, Sally, Jane$}

$$\chi_{intelligent}(x) = \begin{cases} 0.9 & \text{x=John} \\ 0.8 & \text{x=Fred} \\ 0.8 & \text{x=Sally} \\ 0.3 & \text{x=Jane} \end{cases}$$

$$\chi_{rich}(x) = \begin{cases} 1.0 & \text{x=John} \\ 0.9 & \text{x=Fred} \\ 0.7 & \text{x=Sally} \\ 0.1 & \text{x=Jane} \end{cases}$$

"All intelligent persons are rich."

$\forall x(intelligent(x) \rightarrow rich(x))$

Let z=$intelligent \rightarrow rich$

$$\chi_z(x) = \begin{cases} 1.0 & \text{x=John} \\ 1.0 & \text{x=Fred} \\ 0.9 & \text{x=Sally} \\ 0.8 & \text{x=Jane} \end{cases}$$

That is, a certainty factor of 0.8 is assigned to the above statement.

Dempster/Shafer Belief Functions

Suppose Θ is a finite set, and let 2^{Θ} denote the set of all subsets of Θ. Suppose the function Bel: $2^{\Theta} \to [0,1]$ satisfies the following conditions:

(1) Bel($\emptyset$)=0

(2) Bel(Θ)=1

(3) $Bel(A_1 \cup ... \cup A_n) \geq$

$$\sum_i Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + ... - ... + (-1)^{n+1} Bel(A_1 \cap ... \cap A_n)$$

then Bel is called *a belief function over* Θ.

remark: Belief function are models to approximate our patial belief in certain propositions.

Bayesian Belief Functions

(1) Bel($\emptyset$)=0

(2) Bel(Θ)=1

(3) If $A \cap B = \emptyset$ then $Bel(A \cup B) = Bel(A) + Bel(B)$ — *Bayes' rule of additivity.*

(4) If Bel(A)>0 then $Bel(B|A) = \frac{Bel(A \cap B)}{Bel(A)}$

Remarks:

- Bayesian Belief functions are a special case of Dempster/Shafer belief functions.
- In the case of a Bayesian belief function, it is impossible to assign belief to A without assigning the remaining belief to not(A), because Bayes' rule of additivity is assumed to hold. That is, it cannot be expressed that let's say mildly positive evidence for A has been observed and nothing concerning not(A) has been observed. Bayes' approach has no representation for **ignorance**.
- Dempster's Rule of Combination is the conterpart of Bayes' theorem in Bayesian approaches.

Basic Probability Assignment Functions

Example

Let $\Theta = \{A, B, C\}$, where

- A="The Dow Jones index will rise by 2 or more percent."
- B="The Dow Jones index will not change dramatically(by more than 2%)."
- C="The Dow Jones index will fall by 2 or more percent."

Obviously, A, B and C are exclusive.

$Bel(\emptyset) = 0$
$Bel(\{A\}) = 0.1$
$Bel(\{B\}) = 0.0$
$Bel(\{C\}) = 0.5$
$Bel(\{A, B\}) = 0.2$
$Bel(\{A, C\}) = 0.6$
$Bel(\{B, C\}) = 0.7$
$Bel(\Theta) = 1.0$

The corresponding probability assignment function m is:

$m(\emptyset) = 0$
$m(\{A\}) = 0.1$
$m(\{B\}) = 0.0$
$m(\{C\}) = 0.5$
$m(\{A, B\}) = 0.1$
$m(\{A, C\}) = 0.0$
$m(\{B, C\}) = 0.2$
$m(\Theta) = 0.1$

Subsets A of Θ for which $m(A) > 0$ are called *focal elements of Bel*.

Dempster's Rule of Combination

let m(A) measure the the total portion of belief, or the total probability mass, that is confined to A yet none of which is confined to any proper subset of A.

Example: Applying Dempster's Rule of Combination

Let $\Theta = \{A, \sim A\}$ and m1 and m2 be defined as follows:

$m_1(A) = 0.7$

$m_1(\sim A) = 0.1$

$m_1(\Theta) = 0.2$

$m_2(A) = 0.4$

$m_2(\sim A) = 0.2$

$m_2(\Theta) = 0.4$

$$m_1 \oplus m_2(A) = \frac{0.4 * 0.7 + 0.7 * 0.4 + 0.4 * 0.2}{1 - 0.1 * 0.4 - 0.2 * 0.7} = \frac{0.64}{0.82}$$

$$m_1 \oplus m_2(\sim A) = \frac{0.1 * 0.2 + 0.1 * 0.4 + 0.2 * 0.2}{0.82} = \frac{0.1}{0.82}$$

$$m_1 \oplus m_2(\Theta) = \frac{0.4 * 0.2}{0.82} = \frac{0.08}{0.82}$$

Properties of $\oplus$

- $\oplus$ is commutative and associative; that is, the order in which evidence is combined doesn't matter.
- $\oplus$ is belief-function preserving: if applied to belief functions the resulting function is a belief function.

Relationship to MYCIN

Combination of Non-Conflicting Evidence

Set

$m_1(A) = a$, $m_2(A) = b$

$m_1(\sim A) = 0$, $m_2(\sim A) = 0$

$m_1(\Theta) = 1 - a$, $m_2(\Theta) = 1 - b$

$$m_1 \oplus m_2(A) = \frac{a * b + a * (1 - b) + b * (1 - a)}{1 - 0 - 0} = a + b - a * b$$

That is, MYCIN's combination functions can be regarded as a special case of Dempster's rule of combination.

Combination of Conflicting Evidence

$m_1(A) = a$, $m_2(A) = 0$

$m_1(\sim A) = 0$, $m_2(\sim A) = b$

$m_1(\Theta) = 1 - a$, $m_2(\Theta) = 1 - b$

$$m_1 \oplus m_2(A) = \frac{a - a * b}{1 - a * b}$$

$$m_1 \oplus m_2(\sim A) = \frac{b - a * b}{1 - a * b}$$

$$m_1 \oplus m_2(\Theta) = \frac{(1 - a) * (1 - b)}{1 - a * b}$$

Remarks

- Obviously, by applying Dempster's rule of combination we have received a different combination function.
- The function is more sophisticated compared to MYCIN's combination function, which is not very suprising because because it uses more information; it uses a two-valued approach in contrast to MYCIN which uses a single-valued approach: we have to store two values m(A) and m($\sim$A) in contrast to MYCIN which uses either one of the two.

Dow-Jones Example continued

Independent Evidence has been received from another knowledge source expressed by m_1, which has the following focal elements:

$m_1(\{C\}) = 0.2$
$m_1(\{B, C\}) = 0.5$
$m_1(\Theta) = 0.3$

After combining the evidence expressed by m_1 and m, we receive:

$$m \oplus m_1(\{C\}) = \frac{0.2 * (0.5 + 0.2 + 0.1) + 0.5 * 0.5 + 0.3 * 0.5}{1 - (0.2 * (0.1 + 0.1) + 0.5 * (0.1))} = \frac{0.56}{0.91}$$

$$m \oplus m_1(\{B, C\}) = \frac{0.5 * (0.2 + 0.1) + 0.3 * 0.2}{0.91} = \frac{0.2}{0.91}$$

$$m \oplus m_1(\{A\}) = \frac{0 + 0 + 0.3 * 0.1}{0.91} = \frac{0.03}{0.91}$$

remark: The demominator of Dempster's rule enforces that contradictory evidence(originally assigned to $m_1 \oplus m_2(\emptyset)$) is equally distributed over the focal elements of $m_1 \oplus m_2$ guranteeing that $m_1 \oplus m_2(\emptyset)=0$ holds and that the sum over all focal elements of $m_1 \oplus m_2$ is 1.

The Interval Approach

In this approach the belief that a certain proposition P is true is measured by assigning an interval $[a\ b]$ to P, expressing the following semantics:

(1) The probability that P is true is at least a. The *confirmation* of P is a: $\text{conf}[a\ b] = a$
(2) The probability that P is false is at least $(1 - b)$. The *disconfirmation* of P is $(1 - b)$: $\text{disconf}([a\ b]) = 1 - b$.
(3) The *uncertainty* of our belief concerning P is measured by $(b - a)$: $unc([a\ b]) = b - a$.
(4) The *certainty* of our belief concerning P is measured by $1 - (b - a)$: $cert([a\ b]) = 1 - unc([a\ b])$.
(5) The *mean value* of our belief concerning P is measured using:
$mv([a\ b]) = \frac{a+b}{2}$.

For example, if we assign an interval $[40\ 99]$ to P we express the following: P is confirmed with 40%. The disconfirmation of P is 1%. That is, 40% of the probability is assigned to P, 1% pf the probability is assigned to (not P); it is unknown how the remaining probability(59%) is distributed: we don't know how much of this probability is assigned to P and how much is assigned to (not P); the uncertainty is 59%. A special case is the interval $[0\ 1]$; it expresses the fact that we know nothing about a proposition P[1]; the confirmation and the disconfirmation of P is 0%. The uncertainty is 100%.

---

[1] That is, "unknown" can be directly be represented in the interval approach.

2. Computing the certainty factors of the left-hand side

We distinguish the following 5 cases for computing $P(A_1... \wedge ... \wedge A_n)$ assuming the certainty of A_i is $[a_i \ b_i]$ for i=1,n.

case1: correlation=1

The A_i are maximally overlapping(best case)

$$A_1 \ [a_1 \ b_1]$$

$$...$$

$$\frac{A_n[a_n \ b_n]}{A_1 \wedge ... \wedge A_n \ [max(a_1,...,a_n), max(b_1,...,b_n)]}$$

$$P(A_1... \wedge ... \wedge A_n) = min\{P(A_1),...,P(A_n)\}$$

case2: correlation=0

The A_i are statistically independent($\approx$ average case) — the A_i are uncorrelated.

$$A_1 \ [a_1 b_1]$$

$$...$$

$$\frac{A_n[a_n b_n]}{A_1 \wedge ... \wedge A_n \ [a_1 * ... * a_n, b_1 * ... * b_n]}$$

case3: correlation=-1

The A_i are maximally disjoint(worst case) — high negative correlation between the A_i.

Let $\psi(c_1,...,c_n) = max(0, 1 - (1 - c_1) - ... - (1 - c_n))$

$$A_1 \ [a_1 b_1]$$

$$A_1 \ [a_1 \ b_1]$$

$$...$$

$$A_n \ [a_n \ b_n]$$

$$\frac{}{A_1 \wedge ... \wedge A_n \ [\psi(a_1,...,a_n), \psi(b_1,...,b_n)]}$$

case4: $0 < correlation < 1$

Interpolate using the results received for case1 and case2.

case5: $-1 < correlation < 0$

Interpolate using the results for case2 and case3.

For example, assuming correlation=0 we will receive for:

$$(and \ (0.4 \ 0.9)(0.8 \ 0.9)) = (0.32 \ 0.81)$$

Assuming correlation=1, we would receive:

$$(and \ (0.4 \ 0.9)(0.8 \ 0.9)) = (0.4 \ 0.9)$$

The negation of intervals can easily be computed using:

$$not(I) = not((a \ b)) = (1 - b \ \ 1 - a)$$

Finally, the corresponding formulas for the disjunction can easily be derived using the formulas for the conjunction and:

$$P(A \lor B) = P(A) + P(B) - P(A \land B)$$

Another special problem arises when the left-hand side of a rules refers to predicates that are not stored in the knowledge base. Two different interpretations make sense in this case:
1) *closed world assumption*: P is assumed to be false([0 0]).
2) *open world assumption*: P is assumed to be undefined([0 1]).

3. Modus Ponens Generating Functions

The next problem is, how can the positive/negative evidence a rule

```
(R1 (if E) (then (infer H with [c d])) )
```

R1 provides concerning a predicate H be computed, if the certainty factor of the rule's left-hand side has already been computed as [a b]? Schematically:

$$if \ \ E \ \ then \ \ H \ \ [c \ d]$$
$$E \ \ [a \ b]$$

$$\overline{What \ is \ the \ rule's \ contribution \ (?1 \ ?2) \ concerning \ H?}$$

Functions that compute (?1 ?2) are called *modus ponens generating functions*.

The assignment of an interval [c d] to the rule's right hand side predicate H can be interpreted in two ways:

1) $P(H|E)$ is at least c and at most d
2) $P(E \longrightarrow H)$ is at least c and at most d

3.1 Interpretation as a conditional probability

general assumption: if E then H [c d] expresses: *the probability of $P(H|E)$ is at least c and at most d.*

Using this interpretation, we can derive the following modus ponens generating function:

Using
if E then H [c d]
E [a b]
$\frac{a+b}{2} > \theta \wedge (b - a) < \psi^6$
we infer:
H [?1 ?2]
With

$$?1 = min(c * a + (1 - d) * (1 - a), c * b + (1 - d) * (1 - b))$$

$$?2 = min(1, max(d * a + (1 - c) * (1 - a), d * b + (1 - c) * (1 - b)))$$

reasons: Let E' denote the observations that cause to suspect that E is true and let $\bar{E}$ denote the a posteri probability of E infered taking into account the additional observations E'. Then, the above formula can be derived as follows:

$$P(H|E') = P(H \wedge E|E') + P(H \wedge not(E)|E') =$$

Making the "reasonable" assumption that "if we know that E is present(absent), then the observations E' relevant to E provide no further information about H" – that is, $P(H|E \wedge E') = P(H|E)$ and $P(H|E \wedge not(E)) = P(H|(not(E))$ – we receive:

$$= P(E|E') * P(H|E' \wedge E) + P(not(E)|E') * P(H|not(E) \wedge E') =$$

$$P(H|E) * P(\bar{E}) + P(H|not(E)) * P(not(\bar{E}))$$

Unfortunately, $P(H|not(E))$ is unknown — we only can say that $P(H|E) > P(H|not(E))$ if E and H are positively correlated and $P(H|E) < P(H|not(E))$ if E and H are negatively correlated. Having no other clues concerning $P(H|not(E))$, we pragmatically set $P(H|not(E))$ to $P(H|E)$. The magnitude of the error we get by doing this is dependent on $(1-P(\bar{E}))$ and on $|correlation(E, H)|$. In order to keep the error small, rules will only be applied, if there is some positive evidence that E is true (for example we set $\theta = 0.55$ and $\psi = 0.85$).

Additionally taking into account that $P(\bar{E})$ is given as an interval [a b], $I_{\bar{H}}$ has to be computed so that the lower bound becomes minimal and the upper bound becomes maximal (varying possible values x with $a \leq x \leq b$ of $I_{\bar{E}}$) — or we can pragmatically set $P(\bar{E})$ to $\frac{a+b}{2}$). If we use the first approach, additionally taking into account that $c \leq P(H|E) \leq d$ holds, we receive for H's (a posterori interval) [?1 ?2], decribed above.

---

6 The logic behind this is that we only have knowledge about the relationship of E and H, but not of not(E) and H

3.2 Interpretation as a probability of a disjunction

general idea: In contrast to the interpretation given before, this approach interprets *if E then H [c d]* as: *The probability of $P(E \longrightarrow H)$ is at least c and at most d.* That is, the approach assigns an uncertainty interval to:

$$P(A \longrightarrow B) = P(\neg A \vee B) = 1 - P(A) + P(B) - P(\neg A \wedge B)$$

Again, we have to distinguish at least 3 cases depending on how the conjunction in the above formula is computed.

case1: Use the Fuzzy Logic Or-function[7]:
Using:
If E then H [c d]
E [a b]
$(1 - a) \leq d$
infer:
if $(1 - a) < c$
 then H [c d]
 else H [0 d]

case2: Use Conventional Probability Theory[8]:
Using:
if E then H [c d]
E [a b]
$(a \neq 0 \wedge (c + a \geq 1) \wedge (b + d \geq 1))$
infer:
H $[\frac{c+a-1}{a} \quad \frac{b+d-1}{b}]$

case 3: Assuming maximal disjointness[9]:
Using:
if E then H [c d]
E [a b]
$(a + c) \geq 1$
infer:
H [c+a-1 b+d-1]

---

[7] Correlation=1
[8] Correlation=0
[9] Correlation=-1

Proofs of the 3 Modus Ponens Generating Functions

case 1:

$max(1 - P(E), P(H)) \geq c$

Depending on which of the two is smaller we receive:

if $((1 - P(E)) < c)$ then $P(H) \geq c$

else "nothing can be infered about P(H)"

$max(1 - P(E), P(H)) \leq d$

Distinguishing the two cases we can infer:

if $((1 - P(E)) \leq d)$ then $P(H) \leq d$

else "we have run into an inconsistency"

Unfortunately, the exact value of P(E) is not known; we only know that $a < P(E) < b$ holds. We have to choose P(E) so that the lower bound of H becomes minimal and the upper bound becomes maximal. Obviously, this can be achieved by setting P(E) to a, because this maximizes the probability that the else-part of the two if-then-else-statements is executed. By integrating the two if-then-else-statement and by setting P(E) to a, we receive the result.

case 2:

$$P(H) \geq \frac{c + P(E) - 1}{P(E)} = 1 - \frac{1 - c}{P(E)}$$

$$P(H) \leq \frac{d + P(E) - 1}{P(E)} = 1 - \frac{1 - d}{P(E)}$$

Both expressions increase monotonously depending on P(E). Therefore, we set P(E)=a for the lower bound and P(E)=b for the upper bound. After incorperating consistency considerations, we receive the result.

case 3:

$1 - P(E) + P(H) \geq c$

$P(H) \geq c + P(E) - 1$

Furthermore:

$P(H) \leq d + P(E) - 1$

Taking the worst case(P(E)=a) for the lower bound and the best case(P(E)=b) for the upper bound, we receive:

$$[\text{a+c-1} \quad \text{b+d-1}]$$

4. Combining evidence received from several rules

Several rules can provide positive or negative evidence concerning a predicate P. Assuming that n intervals $I_1, ..., I_n$ have been infered using the methods discussed in chapters 3 and 4, the overall certainty factor of P has to be computed combining the evidence expressed in the intervals $I_1, ..., I_n$. Schematically:

$$R_1 \; provided \; evidence \; I_1 = (a_1 \; b_1) \; for \; predicate \; P$$

$$.............$$

$$\frac{R_n \; provided \; evidence \; I_n = (a_n \; b_n) \; for \; predicate \; P}{Overall \; evidence \; I = (?1 \; ?2) \; for \; P}$$

That is, we are looking for functions *combine* that give a reasonable estimation of the certainty I of P:

$$I = combine(I_1, combine(I_2, ..., combine(I_{n-1}, I_n)...))$$

5.1 Requirements for Good Combination Functions

Unfortunately, there is an unlimited number of possible combination functions *combine*. This brings up the question, which combination function is the best one. There is no unique answer to this question — reasoning is performed in different contexts. However, it is worth while to define some general requirements of good combination functions that are capable to facilitate the selection of candidate functions:

(1) More certain clues should get a higher weight than clues that are less confirmed. For example, the meanvalue of $combine([0.5 \; 0.9], [0.40 \; 0.41])$ should be relatively close to 0.41, because the second interval is 40 times less uncertain than the first one.

(2) The addition of very uncertain new evidence should not at all or only slightly affect the overall evidence concerning a predicate P. Especially, $combine([a \; b], [0 \; 1]) = [a \; b]$ must hold.

(3) A higher number of clues (with a similar uncertainty) should lead to a higher certainty of our estimates. In other words, the more evidence we have concerning a predicate of interest P, the more certain is our prediction — the more clues we have analyzed the more certain is the result. New evidence should never increase the uncertainty:
$cert(combine([a \; b], [c \; d])) \geq max\{cert([a \; b]), cert([c \; d])\}$

(4) The order in which certain evidence is combined should have no effect on the overall result; that is, if we combine evidence in different orders, then the result should be the same:

$$combine(combine([a \; b], [c \; d]), [e \; f]) = combine(combine([a \; b], [e \; f]), [c \; d])$$

Remark: If this assumption is violated combining n intervals in different orders may lead to ambiguities, which is not desirable. Furthermore, if the above condition holds, it is sufficient to store for each proposition the current interval and not the set of

previous intervals, which simplifies the implementation of the inference process a lot. On the other hand, this requirement puts severe restrictions on possible combination functions, because combination functions satisfying this requirement have to be associative.

(5) If we have recieved absolute certainty concerning a proposition, further new evidences should not change our overall perception of the proposition; that is we can stop making further inferences concerning P. Formally:

If $b \neq c$ then $combine([a\ a], [b\ c]) = [a\ a]$

Another important aspect – ignored so far in our discussions – is if the evidence concerning a predicate P has been received by interpreting the same or different knowledge sources.

Let's consider the following two examples: Two medical test T1 and T2 are used to predict the probability of P. Both tests interpret the same knowledge source (for example an ECG). T1 comes up with [70 80] and T2 comes up with [80 90]. In this case, we would expect the probability of P about 80% (we apply some kind of voting mechanism), because both intervals have been received from the same knowledge source. Furthermore, if T1 and T2 come up with more or less the same mean value, we would increase the strength of our belief in this probability.

On the hand, let's assume the same intervals have been received concerning a predicate P by interpreting blood cultures(T1) and the patient's ECG(T2); that is, by interpreting two different knowledge sources: In this case, we tend to give a much higher estimation for the probability of P(let's say 98%), because P has been confirmed by interpreting different (and not only single) knowledge sources.

Obviously, we have to distinguish these two cases in our requirements.

(6a) If rules working on different knowledge sources independently provide confirming positive(negative) and reliable[10] evidence, then the mean value of the combined interval should be significantly greater(less) than the maximum mean value of the intervals to be combined: for example,

$$mv(combine([0.9\ 1], [0.8\ 1])) > 0.95$$
$$mv(combine([0\ 0.1], [0\ 0.2])) < 0.05$$

(6b) If evidence has been received using different rules interpreting the same knowledge source, then the combined mean value should not be significantly greater or less than the weighted average of the mean value of the intervals to be combined — using the uncertainty of the intervals as a weight.

---

[10] That is, the uncertainty of the intervals should not be high, let's say less than 0.5.

5.2 Two Combination Functions and their Properties

Two different combination functions are used in PICASSO:

- A function mscomb that assumes that the evidence has been received independently from different knowledge sources. That is, because of the independence assumption mscomb will increase the overall probability of P, if confirming evidence has been found. mscomb can be derived straight forward by applying Dempster's rule of combination ([SHA 76],[DEM 68] — for technical details see [YU 86],[GIN 84] or [DUB 85]).

- A new, so far unpublished, function sscomb that assumes that the evidence to be combined has been received by interpretating the same knowledge source; that is, the different intervals represent differents "opinions" about the same subject. Therefore, sscomb uses a weighted voting approach, that weigths each interval to be combined by its uncertainty: certain intervals get a higher weight compared to uncertain intervals and confirming evidence will not increase the overall probability. For example, see the difference in:

sscomb([0.9 1],[0.9 1])=(0.925 0.975) and mscomb([0.9 1],[0.9 1])=[0.99 1].

mscomb and sscomb are defined as follows:

$$mscomb([l_1 \ u_1], [l_2 \ u_2]) =$$

$$[\frac{l_1 \times u_2 + l_2 \times u_1 - l_1 \times l_2}{1 - l_1 \times (1 - u_2) - l_2 \times (1 - u_1)} \quad \frac{u_1 \times u_2}{1 - l_1 \times (1 - u_2) - l_2 \times (1 - u_1)}]$$

Let $\tau = unc(I_1) + unc(I_2)$

Then sscomb is defined by the following two equations:

$$mv(sscomb(I_1, I_2)) = \begin{cases} 0.5 & \tau = 2 \\[2mm] \frac{cert(I_1)}{cert(I_1)+cert(I_2)} * mv(I_1) + \frac{cert(I_2)}{cert(I_1)+cert(I_2)} * mv(I_2) & 1 < \tau < 2 \\[2mm] \frac{unc(I_2)}{unc(I_1)+unc(I_2)} * mv(I_1) + \frac{unc(I_1)}{unc(I_1)+unc(I_2)} * mv(I_2) & 0 < \tau \leq 1 \\[2mm] \frac{mv(I_1)+mv(I_2)}{2} & \tau = 0 \end{cases}$$

$$unc(sscomb(I_1, I_2)) = \begin{cases} 0 & if \ \tau = 0 \\[2mm] \frac{unc(I_1)*unc(I_2)}{unc(I_1)+unc(I_2)} & 0 < \tau \leq 1 \\[2mm] unc(I_1) * unc(I_2) & 1 < \tau < 2 \\[2mm] 1 & \tau = 2 \end{cases}$$

Obviously, the lower and upper bound of the combined interval I can easily computed using the mean value and the uncertainty.

The two combination functions have the following properties:

- In both functions the uncertainty of the combined evidence is smaller than the uncertainty of the original intervals: *many clues increase the certainty.*
- mscomb([0 1]),[a b])= sscomb([0 1],[a b])=[a b]: *completely uncertain knowledge should not change the overall evidence.*
- mscomb is associative; sscomb is associative when combining intervals, whose uncertainty is less equal 0.5.
- sscomb combines the intervals assigning a higher weight to more certain intervals; that is, it gives preference to evidence received from rules that have a high degree of confirmation. Mscomb does not always have this property[12].
- mscomb and sscomb are commutative.
- MYCIN's combination function is a special case of mscomb.

In summary, dscomb satisfies the requirements (2),(3),(4) and (6a); sscomb satisfies the requirements (1),(2),(3),(4) when combining certain intervals($\tau \leq 1$), (5) and (6b).

---

[12]   see for example:

dscomb([0.2 0.3],[0.0 0.8])=[0.166 0.250]

dscomb([0.2 0.3],[0.2 0.6])=[0.179 0.231]

dscomb([0.2 0.3],[0.4 0.4])=[0.200 0.200]

5. A small example

```
(r1 (if (and (cloudy-sky)(humid))) (then (infer (rain) with (0.4 0.9))) (corr 0.5))
(r2 (if (high-pressure)) (then (infer (rain) with (0.0 0.7))))
(r3 (if (and (hot)(humid))) (then (infer (rain) with (0.6 1.0))) (corr 1.0))
(r4 (if (not(high-pressure))) (then (infer (rain) with (0.3 1.0))))
```

Assumption: dscomb and the modus ponens g. function of 3.1 is used.

Infering (rain)	case1	case2	case3
cloudy-sky	(0.88 0.90)	(0.60 0.62)	(0.90 0.92)
humid	(0.88 0.90)	(0.58 0.60)	(0.62 0.64)
hot	(0.80 0.82)	(0.90 0.94)	(0.65 0.67)
high-pressure	(0.32 0.34)	(0.80 0.82)	(0.90 0.92)
evidence r1 (rain)	(0.35 0.86)	not fired	(0.28 0.78)
evidence r2 (rain)	not fired	(0.05 0.76)	(0.02 0.73)
evidence r3 (rain)	(0.48 0.89)	(0.35 0.76)	(0.37 0.78)
evidence r4 (rain)	(0.20 0.90)	not fired	not fired
overall evidence (rain)	(0.67 0.84)	(0.32 0.64)	(0.40 0.60)

Interpretation

- In the first case there is a strong suggestion that it will rain(about a 75% chance); in the other two cases a about 50% percent of rain chance was predicted. However, for the second test case this prediction is relatively unreliable: the uncertainty is 0.32.
- Note that in case 3 the rule r2 that provided positive evidence that balanced the negative evidence provided by rule r1 and r3.
- Although, the evidence we got in case 1 from r1, r2 and r4 was relatively uncertain, the presense of 3 independent clues supporting the rain hypothesis resulted in a relatively reliable predication of a 75% chance of rain.