Uncertainty Modelling for Intelligent Systems

Lecture Notes

jerakrs

January 2018

Abstract

This document is the lecture note of *Uncertainty Modelling for Intelligent Systems*. The course code is *EMATM1120* and the unit director is *Jonathan Lawry*.

1 Quantitative Measures of Uncertainty

Certainty vs. Uncertainty

- Certainty: Given a number of inference rules we are then able to deduce other propositions as being certainty true.
- Uncertainty: Our knowledge consists of proposition in which we have some level of belief but about which we are not certain.

Aim: investigate frameworks according to which a rational intelligent agent could reason under uncertainty.

1.1 Inductive Logic Rudolf Carnap

There are N unary predicates $Q_1, Q_2, \ldots, Q_N$ which can be applied to M objects $a_1, a_2, \ldots, a_M$, any object must satisfy exactly one of the formula:

$$\alpha(x) = \pm Q_1(x) \wedge \pm Q_2(x) \wedge \cdots \wedge \pm Q_n(x) \tag{1}$$

and $\pm Q_i(x) = Q_i(x)$ or $\neg Q_i(x)$, then for any event p, have:

$$p = \alpha(a_1) \wedge \alpha(a_2) \wedge .... \wedge \alpha(a_M) \tag{2}$$

1.2 Uncertainty Measures

- W: the set of all possible worlds which is assumed be finite.
- $\mu : 2^W \to [0, 1]$ is a function mapping subsets of W into the interval.

e.g. four student, each bit represent that it is present or not:

$$\mu(\{< 1, 0, 0, 1 >\}) = 1$$

1.3 Random Variables

$X : W \to \Omega$ is a function from W into some measurement domain Ω.

- $\mu(X = x) = \mu(\{w : X(w) = x\})$
- For $S \subseteq \Omega, \mu(X \in S) = \mu(\{w : X(w) \in S\})$

e.g. X is the number of male students present, then:

$$\mu(X = 2) = \mu(\{< 1, 0, 0, 1 >, < 1, 0, 1, 1 >, < 1, 1, 0, 1 >, < 1, 1, 1, 1 >\})$$

1.4 Random Sets

$R : W \to 2^{\Omega}$: measurements which are imprecise so that $R(w) = S$ when $S \subseteq \Omega$.

$$\mu(R = S) = \mu(\{w : R(w) = S\}) \tag{3}$$

e.g. Let R denote the lecturer's estimate of the number of people present:

$$R(< 1, 0, 1, 1 >) = \{3, 4\}$$

1.5 Attribute of Measures

- **Additive**
  - For uncertainty measures: If $A \cap B = \varnothing$ then $\mu(A \cup B) = \mu(A) + \mu(B)$
  - General form: For any A and B, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$
- **Sub-Additive**
  - For uncertainty measures: If $A \cap B = \varnothing$ then $\mu(A \cup B) \leq \mu(A) + \mu(B)$
  - General form: For any A and B, $\mu(A \cup B) \leq \mu(A) + \mu(B) - \mu(A \cap B)$
- **Super-Additive**
  - For uncertainty measures: If $A \cap B = \varnothing$ then $\mu(A \cup B) \geq \mu(A) + \mu(B)$
  - General form: For any A and B, $\mu(A \cup B) \geq \mu(A) + \mu(B) - \mu(A \cap B)$
- **Maxitive**
  - For any A and B, $\mu(A \cup B) = max(\mu(A), \mu(B))$: Special case of sub-additive measure.
- **Minimal**
  - For any A and B, $\mu(A \cup B) = min(\mu(A), \mu(B))$: Special case of super-additive measure.

2 Probability Theory

2.1 Axioms

For probability measures we denote the uncertainty measure $\mu = P$

- **P1**: $P(W) = 1, P(\varnothing) = 0$, W is the all possible world.
- **P2**: $A \cap B = \varnothing$ then $P(A \cup B) = P(A) + P(B)$ to general additivity $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Complement: $P(A^c) = 1 - P(A)$
Proof:

$$\because P1, A \cup A^c = W$$
$$\therefore 1 = P(W) = P(A \cup A^c)$$
$$\because P2, A \cap A^c = \varnothing$$
$$\therefore P(A \cup A^c) = P(A) \cup P(A^c)$$
$$Therefore : P(A) \cup P(A^c) = 1$$

Probability Distributions: $A = \{w_1, ..., w_k\}$ so that $A = \{w_1\} \cup .... \cup \{w_k\}$, then $P(A) = P(w_1) + P(w_2) + ... + P(w_k)$

- For uncertainty measure: if W has n elements, agent needs to specify $2^n - 2$ values ($\mu(W) = 1, \mu(\varnothing) = 0$).
- For probability measure: if W has n elements, agent needs to specify $n-1$ values (Probability distribution must sum to 1).

2.2 Probabilities Distributions

- **Joint Probability**: $P(X_1, X_2, \ldots X_n)$, if X_i has k_i values then specifying the joint distribution requires $(\prod_{i=1}^{n} k_i) - 1$.
- **Marginal Probability**: $P(X_1) = \sum_{x_2 \in \Omega} \cdots \sum_{x_n \in \Omega} P(X_2 = x_2, \ldots, X_n = x_n)$
- **Conditional Distribution**: $P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)}$

Independence

- Random variable X_1 is independent of X_2 if $P(X_1 = x_1 | X_2 = x_2) = P(X_1 = x_1)$
- Random variables $X_1, \ldots, X_n$ are independent if $P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$

The number of values which must be specified in order to define the joint distribution is:

- the independent case requires $\sum_{i=1}^{n}(k_i - 1)$ values.
- the fully dependent case requires $\prod_{i=1}^{n}(k_i) - 1$ values.

Conditional Independence: Let U, V, W be exclusive subsets of $\{X_1, ..., X_n\}$, then the variables in U are said to be conditionally independent of variables in V given the variables in $P(U|V,W) = P(U|W)$.

2.3 Conditional Probability

For $A, B \subseteq W$ the conditional probability of A given B is define by:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4}$$

Notice that if $P(B) = 0$ then $P(A|B)$ is undefined.

Bayes Theorem

$$P(H|E) = \frac{P(H \cap E)}{P(E)} = \frac{P(E|H)P(H)}{P(E)} \tag{5}$$

$P(E|H)$ is called the likelihood, $P(H)$ is called a prior probability, and

$$P(E) = P(E|H)P(H) + P(E|H^c)P(H^c) \tag{6}$$

Laplace's principle of insufficient reason

In the absence of any other information all possible worlds should be assumed to be equally probable, i.e. the probability distribution should be uniform.

Sometimes assuming a uniform prior gives you different answers if you transform the problem:

- Let $W = \{w1, w2, w3\}$ and X be a random variable with values in $\{1, 2\}$
- By insufficient reason, then $p(w1) = p(w2) = p(w3) = 1$ and $P(X = 1) = P(X = 2) = 1$
- However, $P(X = 1) = P(\{w : X(w) = 1\}) = 0$ or $= \frac{1}{3}$ or $= \frac{2}{3}$ or $= 1$

2.4 Betting Justification

Notation: Let w^* denote the true world, so $\chi_A(w^*) = 1$ when A is true and $\chi_A(w^*) = 0$ when A is false.

$$\chi_A : W \to \{0,1\} \begin{cases} 1, w^* \in A \\ 0, w^* \notin A \end{cases} \tag{7}$$

Betting Rules:

- $0 \leq p \leq 1$ and $S > 0$
- **Bet 1**: Gain $S(1-p)$ if A is true and lose Sp if A is false.
- **Bet 2**: Lose $S(1-p)$ if A is false and gain Sp if A is ture.

Gain for bet 1 and bet 2

- gain for bet 1 $= S(1-p)\chi_A(w^*) - Sp(1-\chi_A(w^*)) = S(\chi_A(w^*) - p)$
- gain for bet 2 $= -S(1-p)\chi_A(w^*) + Sp(1-\chi_A(w^*)) = -S(\chi_A(w^*) - p)$

When the expectation of gain for bet 1 and bet 2 is equal, we have:

$$\mu(A)S(1-p) = -(1-\mu(A))S(0-p)$$
$$\mu(A)(1-p) + (\mu(A)-1)p = 0$$
$$\mu(A) = p$$

therefore when $p < \mu$, the agent pick bet 1 and when $p > \mu$, the agent pick bet 2.

Dutch Book Problem: A Dutch book is a sequence of bets the outcome of which is a sure loss no matter what the actual state of the world. The gain from two bets is $S(\chi_A(w^*) - p) + S(\chi_{A^c}(w^*) - p)$ where w^* is the actual state of the world. Simplifying gives

$$S(\chi_A(w^*) - p) + S(\chi_{A^c}(w^*) - p)$$
$$= S(\chi_A(w^*) + \chi_{A^c}(w^*) - 2p) = S(1-2p) < 0$$

since $S > 0$ and $p > \frac{1}{2}$.

2.5 Bayesian Networks

A Bayesian network is a directed graph (V, E) where $V = \{X_1, ..., X_n\}$ enumerated such that $(X_j, X_i) \in E$ only if $j < i$ together with a probability distribution on V satisfying:

$$P(X_i|X_1, ..., X_{i-1}) = P(X_i| \sqcap (X_i)) \tag{8}$$

where $\sqcap(X_i) = \{X_j : (X_j, X_i) \in E\}$ are the parents of X_i.

- Full dependent: for any $1 \leq i \leq N$ and $\sqcap(X_i) = \{X_j | j < i\}$.
- Full independent: for any $1 \leq i \leq N$, $\sqcap(X_i) = \varnothing$.

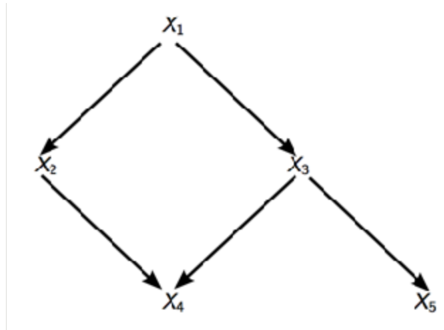

Figure 1: The example of Bayesian Network

PS: The graph not exist cycles since it is illegal for one thing dependent its self.

The Chain Rule:

$$\begin{aligned} &P(X_1, ..., X_n) \\ &\quad = P(X_1)\frac{P(X_1, X_2)}{P(X_1)}\frac{P(X_1, X_2, X_3)}{P(X_1, X_2)}\cdots\frac{P(X_1, ..., X_n)}{P(X_1, \ldots, X_{n-1})} \\ &\quad = P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, ..., X_{n-1}) = \prod_{n=1}^{n} P(X_i| \sqcap (X_i)) \end{aligned} \quad (9)$$

Possible Values: Suppose that $X_1, X_2, \ldots, X_n$ is a binary distribution, the total number of possible values is $\sum_{i=1}^{n} 2^{|\sqcap(x_i)|}$, since $0 \leq |\sqcap(x_i)| \leq i-1$, we have

$$(\sum_{i=1}^{n} 2^0 = n) \leq \sum_{i=1}^{n} 2^{|\sqcap(x_i)|} \leq (\sum_{i=1}^{n} 2^{i-1} = 2^n - 1) \quad (10)$$

which means the possible values satisfy $independent \leq BayesianNetwork \leq fullydependent$.

3 Information and Inference

3.1 Linear Knowledge

A base knowledge can be represented by a set of linear equations on a probability measure P:

$$K = \sum_{i=1}^{n_j} a_{i,j}P(A_{i,j}) = b_j : j = 1, \ldots, m \quad (11)$$

Example of Knowledge: Supposing there are three sections, red(r), blue(b) and green(g). And our knowledge base is the probability of red is twice to probability of blue, $K = \{P(r) = 2P(b)\}$.

- Let denotes $P(r) = p_1, P(b) = p_2, P(g) = p_3$.
- From K we have that $p_1 = 2p_2$ also we know hint $p_1 + p_2 + p_3 = 1 \rightarrow p_3 = 1 - 3p_2$
- The $V(K) \subseteq V$ corresponding to all probability distributions which satisfy K
- $V(K) = \{< 2p_2, p_2, 1 - 3p_2 >: 0 \leq p_2 \leq \frac{1}{3}\}$
- Notice $p_2 \leq \frac{1}{3}$ since we require $3p_2 \leq 1$ and $2p_2 \leq 1$

Recall some basic maths:

- $\log_2 x = \frac{\ln x}{\ln 2}$, let $y = log_a x \Rightarrow a^y = x \Rightarrow lna^y = lnx \Rightarrow y = \frac{\ln x}{\ln a}$
- $\frac{d \ln x}{dx} = \frac{1}{x}$
- $\frac{d(-x \log_2 x)}{dx} = \frac{d(-x \frac{\ln x}{\ln 2})}{dx} = -\frac{1}{\ln 2} \frac{d(x \ln x)}{dx} = -\frac{1}{\ln 2}(\ln x + x\frac{1}{x}) = -\log_2 x - \frac{1}{\ln 2}$

3.2 Information and Entropy

Shannon's Entropy Measure: Let $W = \{w_1, ..., w_n\}$ and $P(w_i) = p_i$ then the information content of this distribution is

$$H := -\sum_{i=1}^{N} p_i log_2(p_i) \tag{12}$$

- H is minimal when $p_j = 1$ and for some $j \in \{1, ..., n\}$ and $p_i = 0$ for all $i \neq j$
- H is maximal when $p_i = \frac{1}{n}$ for $i = \{1, \ldots, n\}$

Proof Maximal Entropy: Let $p_n = 1 - \sum_{i=1}^{n-1} p_i$ and $H = \sum_{i=1}^{n} -p_i log_2 p_i$

$$\frac{dH}{dp_i} = \frac{d(-p_i log_2 p_i)}{dp_i} + \frac{d(-p_n log_2 p_n)}{dp_i} = -log_2 p_i - \frac{1}{\ln 2} + log_2(1 - \sum_{i=1}^{n-1} p_i) + \frac{1}{\ln 2}$$

$$\frac{dH}{dp_i} = -log_2 p_i + log_2(1 - \sum_{i=1}^{n-1} p_i)$$

$$\frac{dH}{dp_i} = 0 \Rightarrow log_2 p_i = log_2 1 - \sum_{i=1}^{n-1} p_i \Rightarrow p_i = p_n$$

$$1 = \sum_{i=1}^{n} p_i \Rightarrow p_i = \frac{1}{n}$$

Center of Mass: Suppose we give every element of $V(K)$ equal probability, then we obtain the probability distribution

$$\hat{p_i} = \frac{\int_{V(K)} p_i dV(K)}{\int_{V(K)} dV(K)} \tag{13}$$

$< \hat{p_1}, \ldots, \hat{p_n} >$ is the centre of mass of $V(K)$.

Example of CM: For $V(K) = \{< p_1, 0.8 - p_1, 0.7 - p_1, p_1 - 0.5 >: 0.5 \leq p_1 \leq 0.7\}$,

$$\hat{p_1} = \int_{V(K)} p_1 dV(K) = \frac{\int_{0.5}^{0.7} p_1 dp_1}{\int_{0.5}^{0.7} dp_1} = \frac{0.12}{0.2} = 0.6$$

4 Dempster-Shafer Theory

4.1 Mass Assignments

Shafer-Dempster belief functions are defined by a probability distirbution on the power set of W.

$$m : 2^w \rightarrow [0,1], m(\varnothing) = 0, \sum_{A \subseteq W} m(A) = 1 \tag{14}$$

Belief and Plausibility

- Belief $Bel : 2^W \rightarrow [0,1]$ is a measure of the total amount of evidence in favor of a proposition

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{15}$$

- Plausibility $Pl : 2^W \rightarrow [0,1]$ is a measure of the total amount of evidence not against a proposition.

$$Pl(A) = \sum_{B : B \cap A \neq \varnothing} m(B) \tag{16}$$

- $\forall A \subseteq W$, $Bel(A) \leq Pl(A)$ and $Pl(A) - Bel(A)$ provides a direct measure of ignorance about the proposition A
- $Bel(A) + Bel(A^c) \leq 1$
- $Pl(A) = 1 - Bel(A^c)$

- Belief is sub-additive: $Bel(A \cup B) \geq Bel(A) + Bel(B)$

$$Bel(A \cup B) = \sum_{C \subseteq A \cup B} m(C) = \sum_{C \subseteq A} m(C) + \sum_{C \subseteq B} m(C) + \sum_{C \subseteq A \cup B, C \nsubseteq A, C \nsubseteq B} m(C)$$

$$\geq \sum_{C \subset A} m(C) + \sum_{C \subset B} m(C) = Bel(A) + Bel(B)$$

- Plausibility is super-additive: $Pl(A \cup B) \leq Pl(A) + Pl(B)$

$$Pl(A \cup B) = \sum_{C \cap (A \cup B) \neq \varnothing} m(C) = \sum_{C \cap A \neq \varnothing} m(C) + \sum_{C \cap B \neq \varnothing} m(C)$$

$$- \sum_{C \cap A \neq \varnothing, C \cap B \neq \varnothing} m(C) \leq \sum_{C \cap A \neq \varnothing} m(C) + \sum_{C \cap B \neq \varnothing} m(C) = Pl(A) + Pl(B)$$

Belief Inversion: If we already have the Belief and we want to know the Mass, we can consider different levels of each elements.

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B) \tag{17}$$

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} (1 - Pl(B^c)) \tag{18}$$

Lower and Upper Probabilities: A lower probability $\underline{P} : 2^W \rightarrow [0, 1]$ and an upper probability $\overline{P} : 2^W \rightarrow [0, 1]$

- $\forall A \subseteq W \quad \underline{P}(A) \leq \overline{P}(A)$
- $\forall A \subseteq W \quad \overline{P}(A^c) = 1 - \underline{P}(A)$

$\mathbb{P}$ is a set of probability measures on 2^W, then $\underline{P}(A) = inf\{P(A) : P \in \mathbb{P}\}$ and $\overline{P}(A) = sup\{P(A) : P \in \mathbb{P}\}$.

4.2 Updating Probabilities

Suppose we have a prior probability measure P and evidence about the true state of the world is given by mass assignment m. To get a single probability measure we could then take the expected value:

$$P(\bullet|m) = \sum_{B \subseteq W} P(\bullet|B) m(B) \tag{19}$$

Example for Updating Probabilities:

$$P(w_1) = 0.5, \; P(w_2) = 0.3, \; P(w_3) = 0.2$$

$$P(w_1|\{w_1, w_2\}) = \frac{P(w_1)}{P(w_1) + P(w_2)} = \frac{5}{8}$$

$$P(w_2|\{w_1, w_2\}) = \frac{P(w_2)}{P(w_1) + P(w_2)} = \frac{3}{8}$$

$$P(w_3|\{w_1, w_2\}) = \frac{P(w_3)}{P(w_1) + P(w_2)} = 0 \text{ } since \ \{w_3\} \cap \{w_1, w_2\} = \varnothing$$

Pignistic Distribution: Take to be the uniform measure so that

$$P_u(A|m) = \sum_{B \subseteq W} \frac{|A \cap B|}{|B|} m(B) \tag{20}$$

where $|A|$ means the number of elements in set A.

4.3 Nested Evidence

Nested hierarchy: There exists $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq W$ such that $\sum_{i=1}^{n} m(F_i) = 1$, $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq W$ are called the **focal sets** of m.

Example for Nested hierarchy: There are $W = \{d_1, d_2, h\}$, suppose all doctors agree that the patient may have disease d_2, a subset of doctors think that he may have disease d_1, some doctors in this subset also believe it is possible that he is healthy.

Possibility and Necessity Measures: For a nested mass assignment the corresponding belief measure is called a **necessity measure** and the plausibility measure is called a **possibility measure**.

- Let F_j be the largest focal set for which $F_j \subseteq A$ then:

$$Bel(A) = Nec(A) = \sum_{B \subseteq A} m(B) = \sum_{i=1}^{j} m(F_i) \tag{21}$$

- Let F_k be the smallest focal set for which $A \cap F_k \neq \varnothing$ then:

$$Pl(A) = Pos(A) = \sum_{B: B \cap A \neq \varnothing} m(B) = \sum_{i=k}^{n} m(F_k) \tag{22}$$

Possibility and Necessity Axioms

- **Pos1**: $Pos(W) = 1, Pos(\varnothing) = 0$
- **Pos2**: $Pos(A \cup B) = max(Pos(A), Pos(B))$
- **Nec1**: $Nec(W) = 1, Nec(\varnothing) = 0$
- **Nec2**: $Nec(A \cap B) = min(Nec(A), Nec(B))$
- **Consequences1**: $\forall A \subseteq W$, $Pos(A) = 1$ or $Pos(A^c) = 1$. Notice that

$$1 = Pos(W) = Pos(A \cup A^c) = max(Pos(A), Pos(A^c))$$

- **Consequences2**: $\forall A \subseteq W$, $Nec(A) = 0$ or $Nec(A^c) = 0$. Notice that
$$0 = Nec(\varnothing) = Nec(A \cap A^c) = min(Nec(A), Nec(A^c))$$
- Either $[Nec(A), Pos(A)] = [x, 1]$ or $[0, y]$ where $0 < x \leq 1$ and $0 \leq y < 1$.

Proof if $Nec(A) > 0$ *then* $Pos(A) = 1$:

- $Pos(A^c) = 1 - Nec(A)$ therefore $Nec(A) > 0 \Rightarrow Pos(A^c) < 1$.
- Now $1 = Pos(W) = Pos(A \cup A^c) = max(Pos(A), Pos(A^c))$.
- Hence, since $Pos(A^c) < 1$ so that $Pos(A) = 1$.

Possibility Distribution: For any $A \subseteq W$, $Pos(A) = max\{Pos(\{w_i\}) : w_i \in A\}$. Let $Pos(\{w_i\}) = \pi(w_i)$ and $max\{\pi(w_i) : w_i \in W\} = 1$.

Possibility Inversion: Suppose $W = \{w_1, \ldots, w_n\}$ ordered, then

$$\begin{gathered} 1 = \pi(w_1) \geq \pi(w_2) \geq \cdots \geq \pi(w_n) \\ m(\{w_1, \ldots, w_n\}) = \pi(w_n) \\ m(\{w_1, \ldots, w_{n-1}\}) = \pi(w_{n-1}) - \pi(w_n) \\ m(\{w_1, \ldots, w_i\}) = \pi(w_i) - \pi(w_{i+1}) \\ m(\{w_1\}) = 1 - \pi(w_2) \end{gathered}$$

4.4 Demspter's Rule of Combination

Suppose we have two mass assignments provided by two independent sources of evidence.

$$\forall A \subseteq W \quad m_1 \bigoplus m_2(A) = \frac{\sum_{(B,C:B \cap C = A)} m_1(B) m_2(C)}{1 - \sum_{(B,C):B \cap C = \varnothing} m_1(B) m_2(C)} \tag{23}$$

Conditional Belief Functions: Suppose that Pl is a plausibility:

$$Pl(A|B) = \sum_{C \subseteq W; C \cap A \neq \varnothing} m \bigoplus m_B(C) = \frac{Pl(A \cap B)}{Pl(B)} \tag{24}$$

Notice that $m \bigoplus m_B(C) = \frac{\sum_{C: C \cap A \neq \varnothing} m(D)}{1 - \sum_{D: D \cap B = \varnothing} m(D)}$

$$Pl(A|B) = \frac{\sum_{D:(D \cap B) \cap A \neq \varnothing} m(D)}{1 - \sum_{D: D \cap B = \varnothing} m(D)} = \frac{\sum_{D:(D \cap B) \cap A \neq \varnothing} m(D)}{\sum_{D: D \cap B \neq \varnothing} m(D)} = \frac{Pl(A \cap B)}{Pl(B)}$$

Also

$$\begin{aligned} Bel(A|B) = 1 - Pl(A^c|B) = 1 - \frac{Pl(A^c \cap B)}{Pl(B)} = \frac{Pl(B) - Pl(A^c \cap B)}{Pl(B)} = \\ \frac{(1 - Bel(B^c)) - (1 - Bel(A \cup B^c))}{1 - Bel(B^c)} = \frac{Bel(A \cup B^c) - Bel(B^c)}{1 - Bel(B^c)} \end{aligned} \tag{25}$$

5 Fuzzy Set Theory

Fuzzy sets are sets to which elements can belong to some partial degree. A (crisp) set $A \subseteq W$ can be characterised by a membership function $\chi_A : W \to \{0, 1\}$ so that

$$\chi_A = \begin{cases} 1: & w \in A \\ 0: & w \notin A \end{cases} \tag{26}$$

The support of a fuzzy set $\tilde{A}$ is given by $Sup(\tilde{A}) = \{w \in W : \chi_{\tilde{A}}(w) > 0\}$ and the fuzzy set is

$$\tilde{A} = \sum_{w \in Supp(\tilde{A})} w/\chi_{\tilde{A}}(w) \tag{27}$$

Sorites Paradox: Consider a man with a lot of hair on his head. One hair is plucked at a time until he is bald. It follows that there must be a hair pluck at which the man switches from being not bald to being bald. However, this is counter intuitive since the difference of one hair cannot induce a category change from not bald to bald. This or an alternative version of sorites.

5.1 T-norm

T-norm are function $T : [0, 1]^2 \to [0, 1]$ which satisfy the following properties:

- **T1**: $\forall x \in [0, 1] \;\; T(x, 1) = x$
- **T2**: $\forall x, y, z \in [0, 1]$, if $y \leq z$ then $T(x, y) \leq T(x, z)$
- **T3**: $\forall x, y \in [0, 1], T(x, y) = T(y, x)$
- **T4**: $\forall x, y, z \in [0, 1], T(x, T(y, z)) = T(T(x, y), z)$

$$\varphi_{(A \cap B) \cap C} = \varphi_{A \cap (B \cap C)} \Rightarrow T(T(A, B), C) = T(A, T(B, C))$$

- **T5**: $\forall x \in [0, 1] \; T(x, x) = x$

Theorem: For any t-norm T, $drastic \leq T \leq min$, where

$$drastic(x, y) = \begin{cases} x & : & y = 1 \\ y & : & x = 1 \\ 0 & : & otherwise \end{cases} \tag{28}$$

- Show that $T \leq min(x, y)$

$$T(x, y) \leq T(x, 1) = x$$

$$T(x, y) = T(y, x) \leq T(y, 1) = y$$

$$T(x, y) \leq min(x, y)$$

- Show that $T \geq drastic(x, y)$

$$T(x, 1) = x$$

$$T(1, y) = T(y, 1) = y$$

$$T(x, y) \geq 0$$

$$T(x, y) \geq drastic$$

Families of T-norms

- **Frank's t-norms**: $T_s(x, y) = log_s[1 + \frac{(s^x-1)(s^y-1)}{s-1}]$ where $s > 0$ and $s \neq 1$
- **Dombi Family**: $T_\lambda(x, y) = \{1 + [(\frac{1}{x} - 1)^\lambda + (\frac{1}{y} - 1)^\lambda]^{\frac{1}{\lambda}}\}^{-1}$, where $\lambda > 0$

5.2 T-conorms

T-conorms are function $S : [0, 1]^2 \to [0, 1]$, it equal to $1 - T(1 - x, 1 - y)$ and satisfy the following properties:

- **S1**: $\forall x \in [0, 1]$ $S(x, 0) = x$
- **S2**: $\forall x, y, z \in [0, 1]$ if $y \leq z$ then $S(x, y) \leq S(x, z)$
- **S3**: $\forall x, y \in [0, 1]$ $S(x, y) = S(y, x)$
- **S4**: $\forall x, y, z \in [0, 1]$ $S(x, S(y, z)) = S(S(x, y), z)$
- **S5**: $\forall x \in [0, 1]$ $S(x, x) = x$

5.3 Fuzzy Set

α**-cuts** provide a way of representing fuzzy sets as a nested sequence of crisp sets, Fuzzy set $\tilde{A}_\alpha = \{w \in W : \chi_{\tilde{A}}(w) \geq \alpha\}$ so $\chi_{\tilde{A}}(w) = \int_{\alpha : w \in \tilde{A}_\alpha} d\alpha$.

Notice that $w \in \tilde{A}_\alpha$ if and only if $\varphi_{\tilde{A}}(w) \geq \alpha$, therefore $\{\alpha : w \in \tilde{A}_\alpha\} = (0, \chi_{\tilde{A}}(w)]$.

Example for Fuzzy Set: There are $\tilde{A} = 5/1 + 6/0.8 + 7/05$ and $\tilde{B} = 1/1 + 2/0.5 + 3/0.2$, use the α-cut method to determine $\frac{\tilde{A}}{\tilde{B}}$.

$\frac{\tilde{A}}{\tilde{B}}$	$\{5, 6, 7\} : (0, 0.5]$	$\{5, 6\} : (0.5, 0.8]$	$\{5\} : (0.8, 1]$
$\{1, 2, 3\} : (0, 0.2]$	$\{5, 6, 7, \frac{5}{2}, 3, \frac{7}{2}, \frac{5}{3}, 2, \frac{7}{3}\} : (0, 0.2]$	$\varnothing$	$\varnothing$
$\{1, 2\} : (0.2, 0.5]$	$\{5, 6, 7, \frac{5}{2}, 3, \frac{7}{2} : (0.2, 0.5]$	$\varnothing$	$\varnothing$
$\{1\} : (0.5, 1]$	$\varnothing$	$\{5, 6\} : (0.5, 0.8]$	$\{5\} : (0.8, 1]$

$$\tilde{A}_\alpha = \begin{cases} \{5, 6, 7\} & : & \alpha \in (0, 0.5] \\ \{5, 6\} & : & \alpha \in (0.5, 0.8] \\ \{5\} & : & \alpha \in (0.8, 1] \end{cases} \ and \ \tilde{B}_\alpha = \begin{cases} \{1, 2, 3\} & : & \alpha \in (0, 0.2] \\ \{1, 2\} & : & \alpha \in (0.2, 0.5] \\ \{1\} & : & \alpha \in (0.5, 1] \end{cases}$$

$$(\frac{\tilde{A}}{\tilde{B}})_\alpha = \begin{cases} \{5, 6, 7, \frac{5}{2}, 3, \frac{7}{2}, \frac{5}{3}, 2, \frac{7}{3}\} & : \quad \alpha \in (0, 0.2] \\ \{5, 6, 7, \frac{5}{2}, 3, \frac{7}{2} & : \quad \alpha \in (0.2, 0.5] \\ \{5, 6\} & : \quad \alpha \in (0.5, 0.8] \\ \{5\} & : \quad \alpha \in (0.8, 1] \end{cases}$$

$$\frac{\tilde{A}}{\tilde{B}} = 5/1 + 6/0.8 + 7/0.5 + \frac{5}{2}/0.5 + 3/0.5 + \frac{7}{2}/0.5 + 5/0.2 + 6/0.2 + 7/0.2$$

6 Model Logic

The idea is to directly incorporate the notions of possibly the case and necessarily the case directly in to a formal logic.

Proposition: $w \models A$ denote that proposition A is true in world w.

$$W(A) = \{w \in W : w \models A\} \tag{29}$$

Well-Formed Formula

- standard connectives: $\neg$ (not), $\wedge$ (and), $\vee$ (or)
- modal operators: $\Box$ (necessarily), $\diamond$ (possibly)
  - $\Box\theta$ means θ is necessarily the case.
  - $\diamond\theta$ means θ is possibly the case.

Semantics of Modal Logic

- $w \models \Box\theta$: for all w' such that $Edge(w, w'), w' \models \theta$
- $w \models \diamond\theta$: if θ is true in every world accessible from w, there exists w' such that $Edge(w, w'), w' \models \theta$
- $W(\Box B) \subseteq W(B) \subseteq W(\diamond B)$
- $\theta \models \varphi$ (θ entails φ or φ follows θ) if and only if $w \models \theta$ implies that $w \models \varphi$
- $\neg\Box\neg\theta \equiv \diamond\theta$ and $\neg\diamond\neg\theta \equiv \Box\theta$

Properties of Relations

- Reflexive: $E(w, w)$ for all $w \in W$
- Symmetric: If $E(w, w')$ then $E(w', w)$
- Transitive: If $E(w_1, w_2)$ and $E(w_2, w_3)$ then $E(w_1, w_3)$
- Dense: If $E(w_1, w_3)$ then there exists w_2 such that $E(w_1, w_2)$ and $E(w_2, w_3)$
- Serial: For all w, there exists w' such that $E(w, w')$

- Identity Relation: $E(w, w')$ if and only if $w = w'$

Probability, Model Logic and DS-Theory: P is a probability measure on 2^W; $R(w)$ is the set of worlds which are reachable from w denoted as $R(w) = \{w' : E(w, w')\}$. Notice that $w \models \theta$ if and only if $R(w) \subseteq W(\theta)$, $w \models \diamond\theta$ if and only if $R(w) \cap W(\theta) \neq \varnothing$.

Mass function: $m(F) = P(\{w : R(w) = F\}) = \sum_{w:R(w)=F} P(w)$

If E is serial then for all w, $R(w) \neq \varnothing$ and hence $m(\varnothing) = 0$, $P(\Box\theta) = Bel(\theta)$ and $P(\diamond\theta) = Pl(\theta)$.

$$P(\Box\theta) = P(W(\Box\theta)) = \sum_{w:R(w)\subseteq W(\theta)} P(w)$$

$$= \sum_{F\subseteq W(\theta)} \sum_{w:R(w)=F} P(w) = \sum_{F\subseteq W(\theta)} m(F) = Bel(W(\theta)) = Bel(\theta)$$

and also $P(\diamond\theta) = P(\neg\Box\neg\theta) = 1 - P(\Box\neg\theta) = 1 - Bel(\neg\theta) = Pl(\theta)$.

Web page Example: Consider the probability distribution $\{w_1, w_2, w_3, w_4, w_5\}$

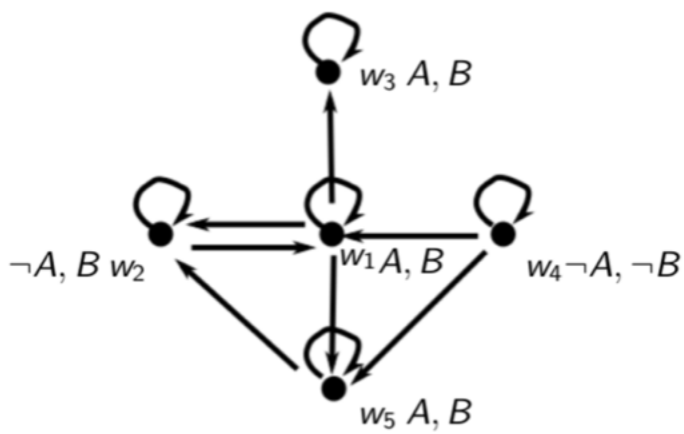

Figure 2: Web Page

$$P(w_1) = 0.1, P(w_2) = 0.2, P(w_3) = 0.2, P(w_4) = 0.4, P(w_5) = 0.1$$

$$R(w_1) = \{w_1, w_2, w_3, w_5\}, R(w_2) = \{w_1, w_2\}, R(w_3) = \{w_3\}$$

$$R(w_4) = \{w_1, w_4, w_5\}, R(w_5) = \{w_2, w_5\}$$

$$W(B) = \{w_1, w_2, w_3, w_5\}$$

$$Bel(B) = m(\{w_1, w_2, w_3, w_5\}) + m(\{w_1, w_2\}) + m(\{w_3\}) + m(\{w_2, w_5\}) = 0.6$$