Le Modèle des Croyances Transférables

Une interprétation de la théorie des fonctions de croyance

David Mercier[1,2]

[1]Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A),

[2]Université d'Artois, IUT Béthune

Séminaire LGI2A, 13 novembre 2007

Imperfection des connaissances

Principales formes d'incertitudes

- L'*incertitude*
  - Relative à la vérité d'une proposition ;
  - Ex : « Je crois que Jean mesure 1,5 mètre »(incertain et précis).
- L'*imprécision*
  - Relative à la nature d'une proposition ;
  - Ex : « Jean mesure entre 1,5 mètre et 2 mètres »(certain et imprécis).
- L'*ambiguïté*
  - Passage graduel d'une catégorie à une autre.
  - Ex : langage courant : « Jean est grand ».
  - Ex : phénomènes naturels : le passage graduel du jour et de la nuit, la maturation d'un fruit.

LGI2A

UNIVERSITÉ D'ARTOIS

Fonctions de croyance

Cadre de représentation des incertitudes

- Un des principaux cadres pour raisonner avec des connaissances imparfaites (imprécises, incertaines, ...), introduit par Dempster (1967) et Shafer (1976).
- Fonctions de croyance généralisent :
  - Mesures de probabilité ;
  - Mesures de possibilité (et sous-ensembles flous).
- Différentes sémantiques de fonctions de croyance :
  - Lower-upper probabilities (modèle de Dempster, Hint model) ;
  - Ensembles aléatoires ;
  - Degrés de croyance (Modèle des Croyances Transférables - MCT).
- Les bases de ce dernier modèle, développé par Ph. Smets, sont présentées ici.

LGI2A

UNIVERSITÉ D'ARTOIS

Fonctions de croyance

Cadre de représentation des incertitudes

- Un des principaux cadres pour raisonner avec des connaissances imparfaites (imprécises, incertaines, ...), introduit par Dempster (1967) et Shafer (1976).
- Fonctions de croyance généralisent :
  - Mesures de probabilité ;
  - Mesures de possibilité (et sous-ensembles flous).
- Différentes sémantiques de fonctions de croyance :
  - Lower-upper probabilities (modèle de Dempster, Hint model) ;
  - Ensembles aléatoires ;
  - Degrés de croyance (Modèle des Croyances Transférables - MCT).
- Les bases de ce dernier modèle, développé par Ph. Smets, sont présentées ici.

LGI2A

UNIVERSITÉ D'ARTOIS

Plan

1. Introduction

2. Le Modèle des Croyances Transférables (MCT)
   - La partie statique (fonctions de croyance)
   - La partie dynamique (combinaison, conditionnement, ...)
   - La partie décisionnelle

3. Conclusion

Plan

LGI2A

UNIVERSITÉ D'ARTOIS

Fonction de Masse

Définition

- $\Omega = \{\omega_1, \ldots, \omega_K\}$: ensemble fini de réponses à une certaine question Q (cadre de discernement).

Définition (fonction de masse)

Une fonction de masse de croyance sur Ω est une application $m : 2^\Omega \to [0, 1]$ t.q.

$$\sum_{A \subseteq \Omega} m(A) = 1.$$

Tout $A \subseteq \Omega$, $m(A) > 0$ est appelé élément focal (EF) de m.

UNIVERSITÉ D'ARTOIS

Fonction de Masse

Interprétation

- La fonction de masse m représente :
  - L'état de connaissance d'un agent rationel Ag à un certain instant t, relativement à Q ;
  - Par extension, un élément d'évidence qui induit un tel état de connaissance.
- $m(A)$: part de croyance allouée à A (et à aucun sous-ensemble strict).
- $m(\Omega)$: degré d'ignorance.
- $m(\emptyset)$: degré de conflit. Joue un rôle d'alarme dans le MCT.

Exemple

The Peter, Paul and Mary Saga (Smets)

- Problème
  - Un juge sait ceci :
    - ★ Big Boss a décidé que M. Jones devait mourir ;
    - ★ 3 tueurs possibles : Peter, Paul, Mary ;
    - ★ Big Boss désigne à pile ou face le sexe du tueur (pièce non truquée) ;
    - ★ Aucune idée sur le choix entre Peter et Paul, dans le cas où un homme est choisi ;
    - ★ M. Jones est tué par un tueur de Big Boss.
  - Comment représenter cette connaissance du juge sur l'auteur du meurtre ?
- Solution dans le MCT
  - k, le tueur, $\in \Omega = \{Peter, Paul, Mary\}$;
  - $m(\{Peter, Paul\}) = 0.5$ et $m(\{Mary\}) = 0.5$.

Fonctions associées

Fonctions de croyance et d'implicabilité

Définition (Fonction de croyance)

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad \forall A \subseteq \Omega$$

Interprétation de $bel(A)$: degré de croyance en A.

Définition (Fonction d'implicabilité)

$$b(A) = bel(A) + m(\emptyset), \quad \forall A \subseteq \Omega$$

LGI2A

UNIVERSITÉ D'ARTOIS

Fonctions associées

Fonction de plausibilité et communalité

Définition (Fonction de plausibilité)

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Omega$$

Propriété : $pl(A) = bel(\Omega) - bel(\overline{A})$.

Définition (Fonction de communalité)

$$q(A) = \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \Omega$$

- Les fonctions *bel*, *b*, *pl*, *q*, *m* sont en correspondance biunivoque
  $\longrightarrow$ représentations équivalentes.

Fonctions associées

Fonction de plausibilité et communalité

Définition (Fonction de plausibilité)

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Omega$$

Propriété : $pl(A) = bel(\Omega) - bel(\overline{A})$.

Définition (Fonction de communalité)

$$q(A) = \sum_{B \supseteq A} m(B), \quad \forall A \subseteq \Omega$$

- Les fonctions *bel*, *b*, *pl*, *q*, *m* sont en correspondance biunivoque $\longrightarrow$ représentations équivalentes.

Exemple

Peter, Paul and Mary Saga

A	$m(A)$	$bel(A)$	$b(A)$	$pl(A)$	$q(A)$
$\emptyset$	0	0	0	0	1
$\{Peter\}$	0	0	0	0.5	0.5
$\{Paul\}$	0	0	0	0.5	0.5
$\{Peter, Paul\}$	0.5	0.5	0.5	0.5	0.5
$\{Mary\}$	0.5	0.5	0.5	0.5	0.5
$\{Peter, Mary\}$	0	0.5	0.5	1	0
$\{Paul, Mary\}$	0	0.5	0.5	1	0
Ω	0	1	1	1	0

Cas particuliers

- Éléments focaux sont des singletons : $m(A) > 0 \Rightarrow \mid A \mid = 1$
  - $bel = pl$, mesure de probabilité ;
  - Probabilité = fonction de croyance maximalement précise.
- Éléments focaux sont emboites $A_1 \subseteq \ldots \subseteq A_n$
  - $pl(A \cup B) = max(pl(A), pl(B))$: pl = mesure de possibilité ;
  - $bel(A \cap B) = min(bel(A), bel(B))$: bel = mesure de nécessité duale ;
  - Mesure de possibilité = fonction de plausibilité consonante (absence de conflit interne).
- Fonctions de croyance englobent mesures de probabilité et de possibilité.

Plan

1. Introduction
2. Le Modèle des Croyances Transférables (MCT)
   - La partie statique (fonctions de croyance)
   - La partie dynamique (combinaison, conditionnement, ...)
   - La partie décisionnelle
3. Conclusion

LGI2A

UNIVERSITÉ D'ARTOIS

Combinaison conjonctive

- Soient deux fonctions de masse m_1 et m_2 issues de deux sources d'informations fiables et distinctes.
- Somme conjonctive :

$$m_{1 \cap\!\!\!\bigcirc 2}(A) = m_1 \cap\!\!\!\!\bigcirc\, m_2(A) = \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega$$

$$q_{1 \cap\!\!\!\bigcirc 2} = q_1 . q_2$$

- Propriétés :
  - commutative et associative ($\rightarrow$ ordre de combinaison des sources n'a pas d'importance) ;
  - élément neutre : ignorance totale $m(\Omega) = 1$, fonction de masse vide noté m_Ω ;
  - non idempotente.

LGI2A

UNIVERSITÉ D'ARTOIS

Degré de conflit

- Degré de conflit $m_{1 \bigcirc\!\!\!\!\cap 2}(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$.
- Joue un rôle d'alarme dans le MCT.
- Plusieurs raisons possibles :
  - L'une des hypothèses de la combinaison conjonctive n'est pas satisfaite (sources non fiables ou non distinctes) ;
  - La vérité est en dehors de Ω (hypothèse du monde ouvert) ;
  - Les sources ne font pas référence au même objet.
- Si $m(\emptyset) = 0$ est imposé. Règle de Dempster = somme conjonctive puis normalisation.

$$m_{1 \oplus 2}(A) = \frac{m_{1 \bigcirc\!\!\!\!\cap 2}(A)}{1 - m_{1 \bigcirc\!\!\!\!\cap 2}(\emptyset)}, \quad \forall A \neq \emptyset$$

Autres combinaisons

- Combinaison disjonctive
  - Soient deux fonctions de masse m_1 et m_2 issues de deux sources d'informations dont l'une au moins est fiable.
  - Somme disjonctive :

$$m_{1 \bigcirc\!\!\!\!\cup 2}(A) = m_1 \bigcirc\!\!\!\!\cup m_2(A) = \sum_{B \cup C = A} m_1(B) m_2(C), \quad \forall A \subseteq \Omega$$

$$b_{1 \bigcirc\!\!\!\!\cup 2} = b_1 . b_2$$

  - Propriétés :
    - commutative et associative ;
    - élément neutre : conflit totale $m(\emptyset) = 1$;
    - non idempotente.
- Moyenne
  - $m = \frac{1}{n} \sum_{i=1}^{n} m_i$.

Exemple

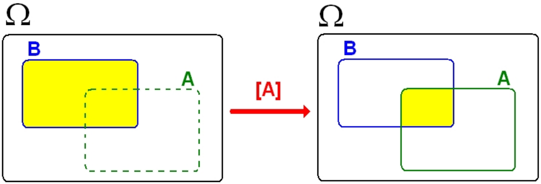
Peter, Paul and Mary Saga

- Un témoin fournit l'information m_t ;
- Combinaison :

A	$m(A)$	$m_t(A)$	$\bigcirc\!\!\!\!\cap(A)$	$\oplus(A)$	$\bigcirc\!\!\!\!\cup(A)$	$moy(A)$
$\emptyset$	0	0	0.45	0	0	0
$\{Peter\}$	0	0	0	0	0	0
$\{Paul\}$	0	0.8	0.40	0.73	0	0.40
$\{Peter, Paul\}$	0.5	0	0.05	0.09	0.40	0.25
$\{Mary\}$	0.5	0.1	0.10	0.18	0.05	0.30
$\{Peter, Mary\}$	0	0	0	0	0	0
$\{Paul, Mary\}$	0	0	0	0	0.40	0
Ω	0	0.1	0	0	0.15	0.05

Règle de conditionnement de Dempster

- Soit m notre fonction de masse représentant notre état de connaissance à t ;
- À $t' > t$, on apprend que la réponse à la question se trouve dans $A \subseteq \Omega$;
- Une nouvelle fonction de masse est obtenue en transférant toute masse $m(B)$ à $B \cap A$;



- D'où l'appellation « modèle des croyances transférables ».

Règle de conditionnement de Dempster

- Soit m notre fonction de masse représentant notre état de connaissance à t ;
- À $t' > t$, on apprend que la réponse à la question se trouve dans $A \subseteq \Omega$;
- Une nouvelle fonction de masse est obtenue en transférant toute masse $m(B)$ à $B \cap A$;
- Cette opération de conditionnement par rapport à A est un cas particulier de la somme conjonctive :

$$m[A] = m \bigcirc\!\!\!\!\cap\, m_A, \quad \text{avec } m_A(A) = 1 \ .$$

Exemple

Peter, Paul and Mary Saga

- Supposons qu'à $t' > t$, le juge apprenne que Peter a un parfait alibi (il n'est pas le coupable, Paul ou Mary est le tueur).
- Révision de la croyance du juge :

A	$m(A)$	$m_{\{Paul,Mary\}}(A)$	$m[Paul, Mary](A)$
$\emptyset$	0	0	0
$\{Peter\}$	0	0	0
$\{Paul\}$	0	0	0.5
$\{Peter, Paul\}$	0.5	0	0
$\{Mary\}$	0.5	0	0.5
$\{Peter, Mary\}$	0	0	0
$\{Paul, Mary\}$	0	1	0
Ω	0	0	0

Notion d'information

- Comment définir le « contenu informationnel »d'une fonction de croyance ?
- Approche ordinale : définition d'un ordre partiel sur l'ensemble des fonctions de croyance.
  - Ex : m_1 est *moins informative* que m_2 ssi

$$\mathrm{pl}_1(A) \geq \mathrm{pl}_2(A) \quad \forall A \subseteq \Omega.$$

- Approche quantitative : définition de « mesures d'incertitude »
  - Ex : la mesure de nonspécificité :

$$N(m) = \sum_{\emptyset \neq A \subseteq \Omega} m(A) \log_2(|A|)$$

LGI2A

UNIVERSITÉ D'ARTOIS

Principe du minimum d'information

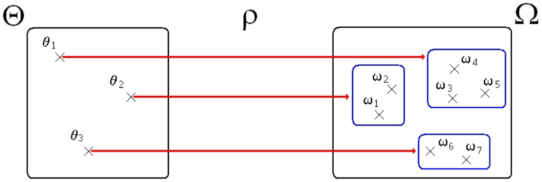
Choisir la fonction de croyance la moins informative (lorsqu'elle existe) parmi l'ensemble des fonctions de croyance compatibles avec les informations disponibles

- Ce principe est un principe de précaution : transfert des masses vers les sous-ensembles les plus larges possibles ;
- Joue le même rôle que principe du maximum d'entropie en probabilités.

Applications :

- extension vide
- déconditionnement.

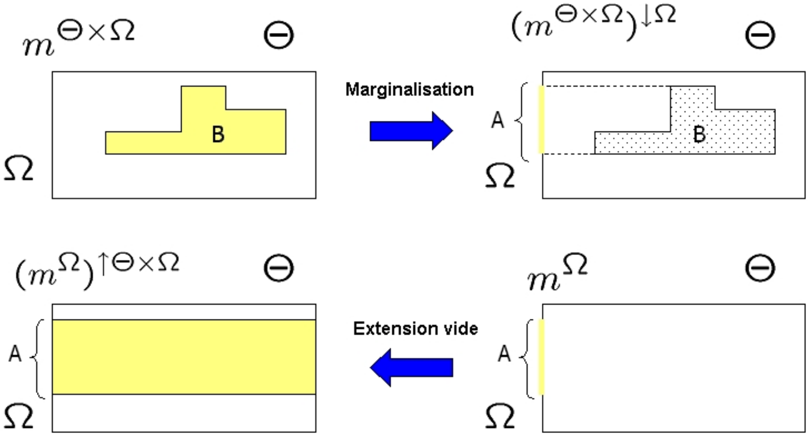
Extension vide



- Soit m^Θ une fonction de masse sur Θ traduisant un certain état de connaissance ;
- Comment exprimer cet état de connaissance dans un réferentiel Ω plus fin ? (Transport de m^Θ dans Ω)
- Solution la moins informative : $m^\Omega(\rho(A)) = m^\Theta(A), \quad \forall A \subseteq \Theta$;
- Définitions :
  - Ω est un raffinement de Θ ;
  - Θ est un grossissement de Ω.

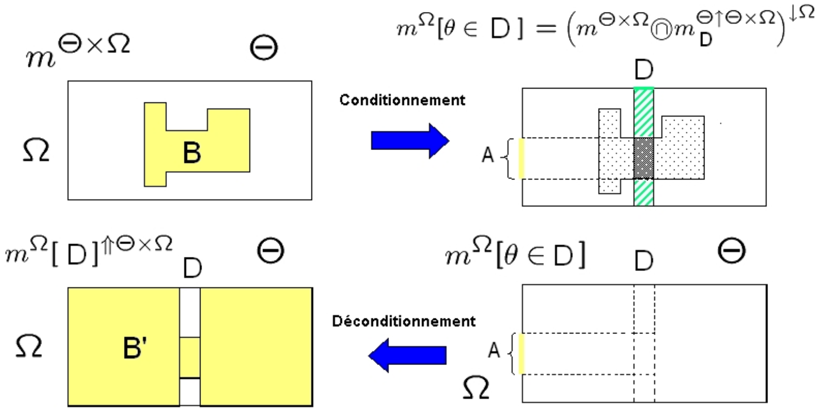
Marginalisation et extension vide

Illustration du transfert d'une masse dans le cas d'un espace produit



Conditionnement et déconditionnement

Illustration du transfert d'une masse dans le cas d'un espace produit



Affaiblissement

- m_S^Ω : information fournie par une source S au regard de Q ;
- La source S est fiable ou non fiable : $\mathcal{R}$=$\{F, NF\}$;
- On suppose :
  - $m^\Omega[\{F\}] = m_S^\Omega$, $m^\Omega[\{NF\}] = m_\Omega$.
  - croyance sur $\mathcal{R}$: $m^\mathcal{R}(\{F\}) = 1 - \alpha$, $m^\mathcal{R}(\mathcal{R}) = \alpha$.
- Combinaison :

$$m^\Omega[m_S^\Omega, m^\mathcal{R}] = \left(m^\Omega[\{F\}]^{\Uparrow \Omega\times\mathcal{R}} \textcircled{\cap} m^{\mathcal{R}\uparrow\Omega\times\mathcal{R}}\right)^{\downarrow\Omega}$$

- Résultat : $m^\Omega[m_S^\Omega, m^\mathcal{R}]$ noté ${}^\alpha m$ = affaiblissement de m_S t.q.

$$\begin{cases} {}^\alpha m(A) & = & (1-\alpha)m_S(A), \quad \forall A \subset \Omega, \\ {}^\alpha m(\Omega) & = & (1-\alpha)m_S(\Omega) + \alpha \ , \end{cases}$$

- Plus simplement : ${}^\alpha m = (1-\alpha)m_S + \alpha\ m_\Omega$.
- Opération donnée par Shafer (1976), preuve issue de Smets.

Exemple

Peter, Paul and Mary Saga

- Supposons :
  - Choix du tueur par Big Boss = correct ;
  - Peter a un alibi ;
  - Fiabilité du témoin = 0.5 .
- Synthèse des croyances du juge :

	m_t	${}^{\alpha}m_t$	$m[Paul, Mary]$	$m[Paul, Mary] \textcircled{\cap} {}^{\alpha}m_t$
$\emptyset$	0	0	0	0.225
$\{Peter\}$	0	0	0	0
$\{Paul\}$	0.8	0.40	0.5	0.475
$\{Peter, Paul\}$	0	0	0	0
$\{Mary\}$	0.1	0.05	0.5	0.300
$\{Peter, Mary\}$	0	0	0	0
$\{Paul, Mary\}$	0	0	0	0
Ω	0.1	0.55	0	0

UNIVERSITÉ D'ARTOIS

Prolongement : affaiblissement contextuel

Prise en compte de connaissances plus fines sur la fiabilité de la source

- m_S^Ω : information fournie par une source S au regard de Q ;
- La source S est fiable ou non fiable : $\mathcal{R}$=$\{F, NF\}$;
- On suppose :
  - $m^\Omega[\{F\}] = m_S^\Omega$, $m^\Omega[\{NF\}] = m_\Omega$;
  - Croyance sur $\mathcal{R}$:
    - $m^\mathcal{R}[\{\omega_k\}](\{F\}) = 1 - \alpha_k$, $m^\mathcal{R}[\{\omega_k\}](\mathcal{R}) = \alpha_k$, $\quad \forall \omega_k \in \Omega$;
    - Fiabilité de la source connue dans chaque contexte $\{\omega_k\}$.
- Combinaison :

$$m^\Omega \left[m_S^\Omega, m^\mathcal{R}[\{\omega_1\}], \ldots, m^\mathcal{R}[\{\omega_K\}] \right]$$

$$= \left(m^\Omega[\{F\}]^{\Uparrow \Omega \times \mathcal{R}} \cap\!\!\!\!\bigcirc\, m^\mathcal{R}[\{\omega_1\}]^{\Uparrow \Omega \times \mathcal{R}} \cap\!\!\!\!\bigcirc \ldots \cap\!\!\!\!\bigcirc\, m^\mathcal{R}[\{\omega_K\}]^{\Uparrow \Omega \times \mathcal{R}} \right)^{\downarrow \Omega}.$$

Plan

1. Introduction
2. Le Modèle des Croyances Transférables (MCT)
   - La partie statique (fonctions de croyance)
   - La partie dynamique (combinaison, conditionnement, ...)
   - La partie décisionnelle
3. Conclusion

Prise de décision

Principes généraux (Théorie de la décision classique)

- Au moment de prendre une décision il faut préciser :
  1. L'ensemble des décisions $\mathcal{D}$ pouvant être prises ;
  2. Les états de la nature considérés = cadre de pari noté Γ.
- Principes de rationalité (Savage, DeGroot) : choisir la décision d minimisant le risque espéré

$$\rho(d) = \sum_{\gamma \in \Gamma} c(d, \gamma) P^{\Gamma}(\{\gamma\})$$

  - $c : \mathcal{D} \times \Gamma \to \mathbb{R}$: une fonction de coût ;
  - $P^{\Gamma} : 2^{\Gamma} \to [0, 1]$: une mesure de probabilité.

LGI2A

UNIVERSITÉ D'ARTOIS

Prise de décision

Dans le MCT

- Dans le MCT, ces principes sont acceptés :
  - Γ = composition de grossissements ou raffinements de Ω ;
  - Transformation pignistique : $m^{\Omega} \rightarrow m^{\Gamma} \rightarrow$ probabilité

$$betP(\{\gamma\}) = \sum_{\{A \subseteq \Gamma, \gamma \in A\}} \frac{m(A)}{\mid A \mid} \times \frac{1}{1 - m(\emptyset)}, \quad \forall \gamma \in \Gamma.$$

Exemple

Peter, Paul and Mary Saga

- Avec $\Gamma = \Omega$ et en omettant les $\{\}$:

	$\emptyset$	Pe	Pa	Pe, Pa	Ma	Pe, M	Pa, M	Pe, Pa, M
m	0	0	0	0.5	0.5	0	0	0
$betP$		0.25	0.25	0.5	0.5	0.75	0.75	1

- Avec $\mathcal{D} = \cup\{d_i\}$:
  - $\rho(d_i) = c(d_i, Pe)betP(\{Pe\}) + c(d_i, Pa)betP(\{Pa\}) + c(d_i, M)betP(\{M\})$.
- Avec $\mathcal{D} = \Omega$, et des coûts 0-1 ($c(\gamma_i, \gamma_j) = 0$ si $\gamma_i = \gamma_j$, 1 sinon) :
  - Décision minisant ρ = Décision de probabilité maximum ;
  - $\rho(Pe) = 0.75 = 1 - betP(\{Pe\})$, $\rho(Pa) = 0.75$, $\rho(M) = 0.5$.

Exemple

Peter, Paul and Mary Saga

- Avec Γ = Ω et en omettant les $\{\}$:

	$\emptyset$	Pe	Pa	Pe, Pa	Ma	Pe, M	Pa, M	Pe, Pa, M
m	0	0	0	0.5	0.5	0	0	0
$betP$		0.25	0.25	0.5	0.5	0.75	0.75	1

- Avec $\mathcal{D}$ = $\cup\{d_i\}$:
  - $\rho(d_i) = c(d_i, Pe)betP(\{Pe\}) + c(d_i, Pa)betP(\{Pa\}) + c(d_i, M)betP(\{M\})$.
- Avec $\mathcal{D}$ = Ω, et des coûts 0-1 ($c(\gamma_i, \gamma_j) = 0$ si $\gamma_i = \gamma_j$, 1 sinon) :
  - Décision minisant ρ = Décision de probabilité maximum ;
  - $\rho(Pe) = 0.75 = 1 - betP(\{Pe\})$, $\rho(Pa) = 0.75$, $\rho(M) = 0.5$.

Exemple

Peter, Paul and Mary Saga

- Avec $\Gamma = \Omega$ et en omettant les $\{\}$:

	$\emptyset$	Pe	Pa	Pe, Pa	Ma	Pe, M	Pa, M	Pe, Pa, M
m	0	0	0	0.5	0.5	0	0	0
$betP$		0.25	0.25	0.5	0.5	0.75	0.75	1

- Avec $\mathcal{D} = \cup\{d_i\}$:
  - $\rho(d_i) = c(d_i, Pe)betP(\{Pe\}) + c(d_i, Pa)betP(\{Pa\}) + c(d_i, M)betP(\{M\})$.
- Avec $\mathcal{D} = \Omega$, et des coûts 0-1 ($c(\gamma_i, \gamma_j) = 0$ si $\gamma_i = \gamma_j$, 1 sinon) :
  - Décision minisant ρ = Décision de probabilité maximum ;
  - $\rho(Pe) = 0.75 = 1 - betP(\{Pe\})$, $\rho(Pa) = 0.75$, $\rho(M) = 0.5$.

LGI2A

UNIVERSITÉ D'ARTOIS

Résumé

Principaux concepts du MCT

- Fonction de croyance = opinion pondérée ; à chaque alternative du monde est associé un nombre entre 0 et 1 ;
- Deux niveaux cognitifs sont distingués dans le raisonnement d'un agent :
  - niveau crédal : seules des fonctions de croyance sont manipulées ;
  - niveau décisionnel : une fonction de probabilité est construite.

Bibliographie

G. Shafer.
A mathematical theory of evidence.
Princeton University Press, 1976.

P. Smets.
http://iridia.ulb.ac.be/~psmets
articles (1994,1998,2005).

T. Denœux.
http://www.hds.utc.fr/~tdenoeux
articles et exposés.

LGI2A
UNIVERSITÉ D'ARTOIS

Fin...

Merci de votre attention