Fusion d'informations dans le cadre de la théorie des fonctions de croyance

Thierry Denœux[1]

[1] Université de Technologie de Compiègne
HEUDIASYC (UMR CNRS 6599)

GDR ISIS
« Gestion de l'incertain et Fusion d'Informations »
9 octobre 2008

Problématique de la fusion d'informations

- Fusion d'informations = combinaison d'informations issues de plusieurs sources en vue de déterminer la valeurs de certaines variables.
- Cela suppose que les sources délivrent une information partielle (imprécise et/ou incertaine) sur les variables d'intérêt.
- Il est donc nécessaire de définir des formalismes suffisamment généraux pour représenter et combiner tous types d'informations susceptibles d'être rencontrés.
- Les problématiques de la fusion de l'information et de la gestion de l'incertain (au sens large) sont donc fortement liées.

Typologie des imperfections de l'information

- Soit X une variable à valeurs dans Ω (domaine, cadre de discernement).
- Un élément d'information sur X peut être représenté sour forme d'un couple (valeur, confiance) :
  - l'élément "valeur" correspond à un sous-ensemble de Ω ;
  - l'élément "confiance" exprime un jugement sur la validité de l'information.
- L'élément d'information est
  - imprécis si "valeur" n'est pas un singleton de Ω ;
  - incertain si l'élément "confiance" exprime un doute sur la véracité de l'information.

Exemple

- Soit X = la température de la pièce
- "Il fait entre 15 et 25 degrés" = ($[15, 25]$,certain) $\longrightarrow$ élément d'information certain et imprécis.
- "Il est plausible qu'il fasse 20 degrés" = (20,plausible) $\longrightarrow$ élément d'information incertain et précis.
- "Il fait probablement entre 15 et 25 degrés" = ($[15, 25]$,probable) $\longrightarrow$ élément d'information incertain et imprécis.

Formalismes classiques

1. Théorie de ensembles
   - Calcul par intervalle, estimation à erreur bornée
   - Modélise bien l'imprécision mais n'exprime aucune notion d'incertitude
   - Peu robuste et excessivement conservative
2. Théorie des probabilités
   - Modélise l'incertitude aléatoire (variabilité dans une population ou au cours des répétitions d'une expérience aléatoire).
   - N'exprime pas de notion d'imprécision.

Théorie des fonctions de croyance

- Introduite par Dempster (1968) et Shafer (1976), puis développée par Smets (Modèle des Croyances Transférables) à partir de 1978.
- Généralise à la fois la théorie des ensembles et la théorie des probabilités :
  - Une fonction de croyance peut être vue comme un ensemble généralisé ou comme une mesure non additive
  - Extension de notions probabilistes (conditionnement, marginalisation) et ensemblistes (intersection, union, inclusion, etc.)
- La combinaison des informations (éléments d'évidence) occupe une place centrale dans la théorie : nombreuses applications en fusion d'informations.

Plan de la présentation

1. Théorie des fonctions de croyance
   - Concepts fondamentaux
   - Fusion d'informations
   - Décomposition canonique et règle prudente
2. Principes de construction de fonctions de croyance
   - Affaiblissement
   - Théorème de Bayes généralisé
   - Fonction de croyance prédictive

Fonction de masse de croyance

- Soit Ω un ensemble fini, domaine d'un variable X.
- Fonction de masse de croyance : $m : 2^\Omega \rightarrow [0, 1]$ telle que

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

- Tout sous-ensemble A de Ω tel que $m(A) > 0$ est appelé ensemble focal.
- Interprétation : m représente
  - Un élément d'évidence relatif à la valeur de X.
  - Un état de connaissance (croyance) induit par cet élément d'évidence.

Cas particuliers

- m peut être vue comme :
  - Une famille d'ensembles pondérés $\{(A_i, m(A_i)), i = 1, \ldots, r\}$.
  - Une distribution de probabilité généralisée (masses distribuées sur 2^Ω et non sur Ω).
- Cas particuliers :
  - $r = 1$: fonction de masse catégorique ($\sim$ ensemble). On note m_A la fonction de masse catégorique d'ensemble focal A.
  - $|A_i| = 1$, $i = 1, \ldots, r$: fonction de masse bayésienne ($\sim$ distribution de probabilité).

Fonctions de croyance et de plausibilité

- Fonction de croyance :

$$bel(A) = \sum_{\substack{B \subseteq A \\ B \not\subseteq \overline{A}}} m(B) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad \forall A \subseteq \Omega$$

(degré de croyance (support) en l'hypothèse "$X \in A$")

- Fonction de plausibilité :

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

(degré maximal de croyance susceptible d'être alloué à A après intégration de nouvelles informations)

- $bel \leq pl$.

Relations entre *m*, *bel* et *pl*

- Relations :

$$bel(A) = pl(\Omega) - pl(\overline{A}), \quad \forall A \subseteq \Omega$$

$$m(A) = \begin{cases} \sum_{\emptyset \neq B \subseteq A} (-1)^{|A|-|B|} bel(B), & A \neq \emptyset \\ 1 - bel(\Omega) & A = \emptyset \end{cases}$$

- *m*, *bel* et *pl* sont donc trois représentations équivalentes d'un même élément d'information.

Relations d'inclusion

- Soient m_1 et m_2 deux fonctions de masse sur Ω.
- En quel sens peut-on dire que m_1 est plus informative que m_2 ?
- Cas particulier :
  - Soient m_A et m_B deux fonctions de masse catégoriques.
  - m_A est plus informative (au sens large) que m_B ssi $A \subseteq B$.
- Généralisation à des fonctions de masse quelconques ?

Inclusion faible

- m_1 plus informative (riche) que m_2 au sens des plausibilités (noté $m_1 \sqsubseteq_{pl} m_2$) si

$$pl_1(A) \leq pl_2(A), \quad \forall A \subseteq \Omega.$$

- Propriétés :
  - Généralise l'inclusion : $m_A \sqsubseteq_{pl} m_B \Leftrightarrow A \subseteq B$.
  - Plus grand élément : m_Ω t.q. $m_\Omega(\Omega) = 1$ (fonction de masse vide).

Inclusion forte

- m_1 est une spécialisation de m_2 (noté $m_1 \sqsubseteq_s m_2$) si m_1 s'obtient à partir de m_2 en répartissant chaque masse $m_2(B)$ entre des sous-ensembles de B :

$$m_1(A) = \sum_{B \subseteq \Omega} S(A,B) m_2(B), \quad \forall A \subseteq \Omega,$$

  avec $S(A,B) =$ proportion de $m_2(B)$ transférée à $A \subseteq B$.
- S : matrice de spécialisation.
- Propriétés :
  - Généralise l'inclusion
  - Plus grand élément : m_Ω.
  - $m_1 \sqsubseteq_s m_2 \Rightarrow m_1 \sqsubseteq_{pl} m_2$.

Principe d'engagement minimal

Définition (Principe d'engagement minimal)

Lorque plusieurs fonctions de croyance sont compatibles avec un ensemble de contraintes, la moins informative (au sens d'un certain ordre) doit être choisie.

- Exemple : On sait seulement que $pl(A) = 0$.
- La fonction de masse la moins informative (au sens de $\sqsubseteq_{pl}$ et $\sqsubseteq_s$) vérifiant cette contrainte et $m_0(\overline{A}) = 1$. Elle vérifie $pl_0(A) = 0$ et $pl_0(B) = 1$ pour tout $B \not\subseteq A$.

Plan de la présentation

1. Théorie des fonctions de croyance
   - Concepts fondamentaux
   - Fusion d'informations
   - Décomposition canonique et règle prudente
2. Principes de construction de fonctions de croyance
   - Affaiblissement
   - Théorème de Bayes généralisé
   - Fonction de croyance prédictive

Opérateurs conjonctifs et disjonctifs

- Soient m_1 et m_2 deux fonctions de masse sur Ω issues de 2 sources.
- Un opérateur de combinaison fusionne m_1 et m_2 en une nouvelle fonction de masse $m_{1*2} = m_1 * m_2$ qui traduit notre nouvel état de connaissance après avoir pris en considération les deux 2 sources.
- L'opérateur $*$ est
  - conjonctif si m_{1*2} est plus informative que m_1 et m_2.
  - disjonctif si m_{1*2} est moins informative que m_1 et m_2.
- Le choix d'un opérateur conjonctif suppose que les sources soient toutes deux fiables.
- Combinaison disjonctive : stratégie prudente (l'une au moins des deux sources est fiable).

Conditionnement

- Soient deux fonctions de masse m et m_B avec $m_B(B) = 1$ pour $B \subset \Omega$.
- Etat de connaissance après fusion conjonctive de m et m_B ? Soit $m[B]$ le résultat de la combinaison.
- Contraintes :
  - $pl[B](\overline{B}) = 0$.
  - $m[B] \sqsubseteq_s m$.
- Principe d'engagement minimal : solution la moins informative au sens de $\sqsubseteq_{pl}$:

$$m[B](A) = \sum_{\{C | C \cap B = A\}} m(C).$$

Propriétés

- Généralisation de l'intersection : $m_A[B] = m_{A\cap B}$.
- Généralisation du conditionnement probabiliste :
  - Si $m(\emptyset) > 0$, soit m^* la fonction de masse normalisée

$$m^*(A) = \frac{m(A)}{1 - m(\emptyset)}.$$

  - Expression du conditionnement normalisé :

$$pl[B]^*(A) = \frac{pl(A \cap B)}{pl(B)}$$

  - Si m est bayésienne, $pl = P$: on retrouve le conditionnement probabiliste.

Plausibilité, communalité

- Interprétation de $pl(A)$:
  - $pl(A) = bel[A](A) = \max_B bel[B](A)$
  - degré maximal de croyance susceptible d'être alloué à A après conditionnement.
- Fonction de communalité : soit $q : 2^\Omega \rightarrow [0, 1]$ définie par $q(A) = m[A](A)$:
  - Masse attachée au plus grand ensemble possible (degré d'ignorance) après conditionnement par A.
  - Autre expression

$$q(A) = \sum_{B \supseteq A} m(B).$$

  - Nouvelle relation d'inclusion : $m_1 \sqsubseteq_q m_2$ si $q_1 \leq q_2$.

Règle de Dempster : hypothèses

- Soit $*$ un opérateur de fusion conjonctif, tel que

$$m_1 * m_2 = S_1 \cdot m_2, \quad m_1 * m_2 = S_2 \cdot m_1,$$

  S_1 et S_2 étant des opérateurs de spécialisation.
- Hypothèses :
  - Indépendance : S_1 ne dépend pas de m_2, S_2 ne dépend pas de m_1.
  - Généralisation du conditionnement : $m_1 * m_B = m[B]$.
  - Commutativité : $m_1 * m_2 = m_2 * m_1$.
- Solution : règle de Dempster.

Règle de Dempster

Définition (Règle de combinaison de Dempster)

$$\forall A \subseteq \Omega, \quad (m_1 \bigcirc\!\!\!\!\cap\, m_2)(A) = \sum_{B \subseteq \Omega} m_1[B](A) m_2(B),$$
$$= \sum_{B \cap C = A} m_1(B) m_2(C).$$

Propriétés :

- commutativité, associativité.
- élément neutre : m_Ω
- $(q_1 \bigcirc\!\!\!\!\cap\, q_2) = q_1 \cdot q_2$
- $(m_1 \bigcirc\!\!\!\!\cap\, m_2)(\emptyset) \geq 0$: degré de conflit.

Règle disjonctive

Définition (Règle disjonctive)

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

Propriétés :

- commutativité, associativité.
- élément neutre : $m_\emptyset$
- Soit $b = bel + m(\emptyset)$ (fonction d'implicabilité). On a :

$$(b_1 \bigcirc\!\!\!\!\cup\, b_2) = b_1 \cdot b_2$$

Choix d'une règle

- Les règles $\textcircled{\cap}$ et $\textcircled{\cup}$ supposent que les sources d'information sont distinctes (la même information n'est pas comptée deux fois). Si cette hypothèse n'est pas vérifiée : règle prudente.
- La règle conjonctive suppose que les deux sources sont fiables : produit une information de plus en plus précise, mais génère du conflit.
- La règle disjonctive suppose que l'une au moins des sources est fiable. Ne génère pas de conflit mais produit une information de moins en moins précise.

Gestion du conflit

- Un conflit important révèle souvent (toujours ?) un problème de modélisation :
  - Crédit excessif accordé à certaines sources ;
  - Hypothèse d'indépendance des sources non vérifiée ;
  - Cadre de discernement non exhaustif, etc...
- Approches :
  - Revenir sur la modélisation de l'information ;
  - Utiliser des règles robustes (résultat cohérent en cas de conflit).

Règle de Dubois et Prade

- Définition :

$$(m_1 *_{DP} m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) + \sum_{\{B \cap C = \emptyset, B \cup C = A\}} m_1(B) m_2(C), \quad \forall A \neq \emptyset$$

$$(m_1 *_{DP} m_2)(\emptyset) = 0.$$

- Propriétés :
  - $m_1 *_{DP} m_2 = m_1 \bigcirc\!\!\!\!\cap\, m_2$ si $(m_1 \bigcirc\!\!\!\!\cap\, m_2)(\emptyset) = 0$,
  - $m_1 *_{DP} m_2 = m_1 \bigcirc\!\!\!\!\cup\, m_2$ si $(m_1 \bigcirc\!\!\!\!\cap\, m_2)(\emptyset) = 1$.
  - Commutative mais non associative.

Plan de la présentation

1. Théorie des fonctions de croyance
   - Concepts fondamentaux
   - Fusion d'informations
   - Décomposition canonique et règle prudente
2. Principes de construction de fonctions de croyance
   - Affaiblissement
   - Théorème de Bayes généralisé
   - Fonction de croyance prédictive

Fonction de masse simple

- Définition : m est une fonction de masse simple si elle est de la forme

$$
\begin{aligned}
m(A) &= 1 - w_A \\
m(\Omega) &= w_A,
\end{aligned}
$$

avec $A \subset \Omega$ et $w_A \in [0,1]$.

- Notation : A^{w_A}.
- Propriété : $A^{w_1} \bigcirc\!\!\!\!\cap A^{w_2} = A^{w_1 w_2}$.
- Une fonction de masse est dite séparable si elle peut s'obtenir comme combinaison de fonctions de masse simples :

$$m = \bigcirc\!\!\!\!\!\cap_{A \subseteq \Omega} A^{w(A)}$$

Retrait d'information

- Soit $m_{12} = m_1 \bigcirc\!\!\!\!\cap\, m_2$. On a $q_{12} = q_1 \cdot q_2$.
- Supposons qu'on s'aperçoive que m_2 a été ajoutée par erreur.
- Si m_2 est non dogmatique ($q(A) > 0, \forall A$), m_1 peut être retrouvée à partir de m_{12} par

$$q_1 = q_{12}/q_2.$$

- On note $m_1 = m_{12} \oslash m_2$.
- Remarque : $m_1 \oslash m_2$ n'est pas toujours une fonction de croyance !

Décomposition canonique

Théorème (Smets, 1995)

Toute fonction de masse non dogmatique ($m(\Omega) > 0$) peut être décomposée de manière unique sous la forme :

$$m = \left(\textcircled{\cap}_{A \subset \Omega} A^{w_C(A)} \right) \oslash \left(\textcircled{\cap}_{A \subset \Omega} A^{w_D(A)} \right)$$

avec $w_C(A) \in]0, 1]$, $w_D(A) \in]0, 1]$ et $\max(w_C(A), w_D(A)) = 1$ pour tout $A \subset \Omega$.

- On peut poser $w = w_C / w_D$.
- La fonction $w : 2^\Omega \setminus \Omega \to \mathbb{R}_+^*$ est une nouvelle représentation équivalente d'une fonction de masse non dogmatique (avec *bel*, *pl*, *q*, *b*).

Propriétés de w

- La fonction w est disponible directement lorsque m est construite par accumulation de fonctions de masses simples (cas fréquent).
- Calcul de w à partir de q :

$$\ln w(A) = -\sum_{B \supseteq A} (-1)^{|B|-|A|} \ln q(B), \quad \forall A \subset \Omega.$$

- Règle de Dempster :

$$w_1 \textcircled{\cap} w_2 = w_1 \cdot w_2.$$

Relation d'inclusion $\sqsubseteq_w$

- Soient m_1 et m_2 deux fonctions de masse non dogmatiques. On pose

$$m_1 \sqsubseteq_w m_2 \text{ ssi } w_1 \leq w_2.$$

- Interprétation $m_1 = m_2 \textcircled{\cap} m$ avec m séparable.
- Propriétés :
  - $m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2, \end{cases}$
  - m_Ω est le seul élément maximal de $\sqsubseteq_w$:

$$m_\Omega \sqsubseteq_w m \Rightarrow m = m_\Omega.$$

Règle prudente
Principe

- Un agent reçoit deux fonctions de masse m_1 et m_2 issues de sources fiables.
- Soit m_{12} la fct de masse représentant son état de connaissance après réception de m_1 et m_2.
- m_{12} doit être plus informative que m_1 et m_2. Formellement : $m_{12} \in \mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2)$, pour un $x \in \{w, s, pl, q\}$, avec $\mathcal{S}_x(m)$= ens. des fcts de masse plus informatives que m au sens de $\sqsubseteq_x$.
- Principe d'engagement minimal : on choisit dans $\mathcal{S}_x(m_1) \cap \mathcal{S}_x(m_2)$ la fonction de masse la moins informative au sens de $\sqsubseteq_x$ (si elle existe).

Règle prudente
Définition

Théorème

Soient m_1 and m_2 deux fcts de masse non dogmatiques. L'élément le moins informatif de $\mathcal{S}_w(m_1) \cap \mathcal{S}_w(m_2)$ au sens de $\sqsubseteq_w$ existe et est unique. Il est défini par

$$w_{1 \textcircled{\wedge} 2}(A) = w_1(A) \wedge w_2(A), \quad \forall A \subset \Omega.$$

Définition (Règle prudente)

$$m_1 \textcircled{\wedge} m_2 = \textcircled{\cap}_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)}.$$

Règle prudente
Définition

Théorème

Soient m_1 and m_2 deux fcts de masse non dogmatiques. L'élément le moins informatif de $\mathcal{S}_w(m_1) \cap \mathcal{S}_w(m_2)$ au sens de $\sqsubseteq_w$ existe et est unique. Il est défini par

$$w_{1 \textcircled{\scriptsize$\wedge$} 2}(A) = w_1(A) \wedge w_2(A), \quad \forall A \subset \Omega.$$

Définition (Règle prudente)

$$m_1 \textcircled{\scriptsize$\wedge$} m_2 = \textcircled{\scriptsize$\cap$}_{A \subset \Omega} A^{w_1(A) \wedge w_2(A)}.$$

Règle prudente
Propriétés

- Commutative, associative
- Idempotente : $\forall m,\ m \owedge m = m$
- Distributivité de $\text{\textcircled{$\cap$}}$ par rapport à $\owedge$:

$$(m_1 \text{\textcircled{$\cap$}} m_2) \owedge (m_1 \text{\textcircled{$\cap$}} m_3) = m_1 \text{\textcircled{$\cap$}} (m_2 \owedge m_3), \forall m_1, m_2, m_3.$$

L'élément d'évidence m_1 n'est compté qu'une fois !

- Absence d'élément neutre, mais $m_\Omega \owedge m = m$ ssi m est séparable.

Extensions

- Il existe une règle disjonctive duale de la règle prudente (règle hardie $\textcircled{\vee}$).
- Quatre règles fondamentales :

Résumé

sources	distinctes	non distinctes
toutes fiables	$\textcircled{\cap}$	$\textcircled{\wedge}$
au moins une fiable	$\textcircled{\cup}$	$\textcircled{\vee}$

- On peut également définir une infinité de règles commutatives et associatives généralisant chacune de ces 4 règles, en se basant sur des normes et des conormes triangulaires.

Plan de la présentation

1. Théorie des fonctions de croyance
   - Concepts fondamentaux
   - Fusion d'informations
   - Décomposition canonique et règle prudente
2. Principes de construction de fonctions de croyance
   - Affaiblissement
   - Théorème de Bayes généralisé
   - Fonction de croyance prédictive

Problème posé

- Une source d'information fournit :
  - une valeur
  - un ensemble de valeurs
  - une distribution de probabilité, etc..
- L'information fournie est :
  - non totalement fiable ou
  - non totalement pertinente.
- Exemples :
  - Défaillance possible d'un capteur :
  - Contexte de la mesure plus ou moins favorable ;
  - Information se rapportant à une situation ou un objet ayant seulement un degré de similarité avec la situation ou l'objet considéré (raisonnement à base de cas).

Modélisation

- Deux cadres de discernement :
  - fiabilité ou pertinence de la source : $\mathcal{F} = \{F, NF\}$.
  - Domaine Ω de la variable X d'intérêt.
- Une source S fournit une fonction de masse m_S^Ω.
- Eléments d'information disponibles :
  - $pl^{\mathcal{F}}(\{NF\}) = \alpha$, $pl^{\mathcal{F}}(\{F\}) = 1$.
  - $m^\Omega[F] = m_S^\Omega$.
  - $m^\Omega[NF] = m_\Omega^\Omega$.
- Information induite sur Ω :

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

Expression de la solution

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

- α est appelé coefficient d'affaiblissement : ${}^{0}m^{\Omega} = m_S^{\Omega}$ et ${}^{1}m^{\Omega} = m_{\Omega}^{\Omega}$.
- Expression équivalente :

$$
{}^{\alpha}m^{\Omega} = m_S^{\Omega} \bigcirc\!\!\!\!\cup\, \emptyset^{\alpha}.
$$

- ${}^{\alpha}m^{\Omega} \sqsupseteq_s m_S^{\Omega}$.

Application en classification

Données du problème

- Soit Ω un ensemble de classes, et

$$\mathcal{L} = \{(\mathbf{x}_i, m_i^\Omega), i = 1, \ldots, n\}$$

  où $\mathbf{x}_i$ est un vecteur d'attributs pour l'objet o_i, et m_i^Ω une fonction de masse relative à la classe de l'objet o_i.
- Soit $\mathbf{x}$ le vecteur d'attributs pour un nouvel individu o à classer.
- Problème : Construire une fonction de masse m^Ω relative à la classe de o.

Application en classification

Solution

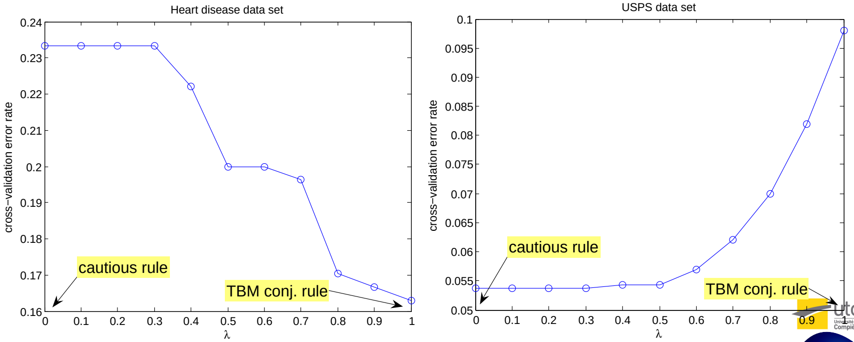
- Hypothèse : soit α_i la plausibilité que *o* et o_i ne soient pas dans la même classe. On suppose $\alpha_i = \phi(d(\mathbf{x}, \mathbf{x}_i))$, où *d* est une distance, et ϕ une fonction croissante de $\mathbb{R}^+$ dans $[0, 1]$.
- Chaque exemple $(\mathbf{x}_i, m_i^\Omega)$ est une source d'information, qui doit être affaiblie avec le coefficient α_i.
- Les *n* fonctions de masse sont ensuite combinées conjonctivement par $\bigcirc\!\!\!\!\cap$ (exemples indépendants) où $\bigcirc\!\!\!\!\wedge$ (exemples non indépendants) :

$$m^\Omega = {}^{\alpha_1}m_1^\Omega * \ldots * {}^{\alpha_n}m_n^\Omega$$

avec $* \in \{\bigcirc\!\!\!\!\cap, \bigcirc\!\!\!\!\wedge, \ldots\}$ (ou tout autre opérateur conjonctif).

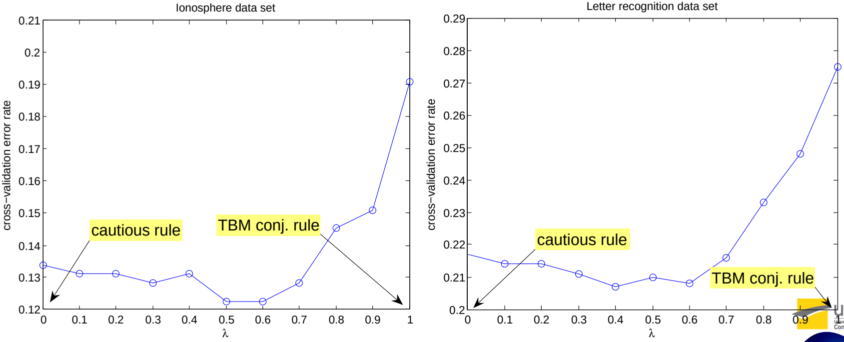
Résultats

Données "Heart disease" and "USPS"



Résultats

Données "Ionosphere" and "Letter recognition"



Généralisation : Affaiblissement contextuel

Modèle

- Modèle plus général de métaconnaissance sur la fiabilité d'une source : la fiabilité dépend du contexte (la vraie valeur de X). Exemples :
  - Fiabilité d'un système de détection dépendant de la nature de la cible
  - Compétence d'un médecin dépendant de la nature de la maladie, etc.
- Modèle :
  - Soit $\Theta = \{\theta_1, \ldots, \theta_L\}$ une partition de Ω.
  - $pl^{\mathcal{F}}[\theta_\ell](\{NF\}) = \alpha_\ell$, $pl^{\mathcal{F}}[\theta_\ell](\{F\}) = 1$, $\ell = 1, \ldots, L$.
  - $m^\Omega[F] = m^\Omega_S$.
  - $m^\Omega[NF] = m^\Omega_\Omega$.

Généralisation : Affaiblissement contextuel
Solution

- Déconditionnement et combinaison de tous les éléments d'information :

$$^{\boldsymbol{\alpha}}m^{\Omega} = \left(\textcircled{\cap}_{\ell=1}^{L} m^{\mathcal{F}}[\theta_\ell]^{\Uparrow \Omega \times \mathcal{F}} \textcircled{\cap} m^{\Omega}[F]^{\Uparrow \Omega \times \mathcal{F}} \right)^{\downarrow \Omega},$$

$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)$.

- Solution :

$$^{\boldsymbol{\alpha}}m^{\Omega} = m_S^{\Omega} \textcircled{\cup} m_1^{\Omega} \textcircled{\cup} \ldots \textcircled{\cup} m_L^{\Omega}$$

avec $m_\ell^{\Omega}(\emptyset) = 1 - \alpha_\ell$, $m_\ell^{\Omega}(\theta_\ell) = \alpha_\ell$, $\ell = 1, \ldots, L$.

- On retrouve l'affaiblissement classique quand $\Theta = \{\Omega\}$.
- Possibilité d'apprentissage des α_ℓ à partir de données étiquetées.

Plan de la présentation

1. Théorie des fonctions de croyance
   - Concepts fondamentaux
   - Fusion d'informations
   - Décomposition canonique et règle prudente
2. Principes de construction de fonctions de croyance
   - Affaiblissement
   - Théorème de Bayes généralisé
   - Fonction de croyance prédictive

Théorème de Bayes généralisé (Smets, 1978)

Problème posé

- Deux variables $X \in \Omega_X$ et $\theta \in \Theta = \{\theta_1, \ldots, \theta_K\}$.
- Typiquement :
  - X est observable (résultat de mesure),
  - θ est non observable (classe, paramètre inconnu à estimer).
- On sait calculer $pl^{\Omega_X}[\theta_k](\{x\}) = pl_k(x)$, $\forall x, k$.
- On n'a pas d'information a priori sur θ : $m^{\Theta}(\Theta) = 1$.
- On observe $X = x$. Fonction de croyance sur Θ ?

Théorème de Bayes généralisé (Smets, 1978)
Solution et propriétés

- Solution (conséquence du principe d'engagement minimal) :

$$m^{\Theta}[x] = \textcircled{\cap}_{k=1}^{K} \overline{\{\theta_k\}}^{pl_k(x)}.$$

- Propriété 1 : Généralisation du théorème de Bayes lorsque $pl_k(x) = P(x|\theta_k)$ (information probabiliste) et combinaison avec une masse Bayésienne sur Θ (a priori).
- Propriété 2 : Si X et Y sont cognitivement indépendantes conditionnellement à θ :

$$pl_k(x,y) = pl_k(x) pl_k(y), \quad \forall k$$

Alors

$$m^{\Theta}[x,y] = \textcircled{\cap}_{k=1}^{K} \overline{\{\theta_k\}}^{pl_k(x,y)} = m^{\Theta}[x] \textcircled{\cap} m^{\Theta}[y].$$

Théorème de Bayes généralisé

Mise en œuvre

- Exemple : Soit X_j un vecteur d'attributs issus d'un capteur j, et $f_k(x_j)$ la densité conditionnelle (estimée) de X_j dans la classe θ_k.
- On peut poser (Appriou, 1991) :

$$pl_k(x_j) = \alpha_{jk} + (1 - \alpha_{jk})\rho_j f(x_j|\theta_k),$$

avec

- ρ_j : coefficient de normalisation
- α_{jk} : coefficient d'affaiblissement traduisant la méconnaissance de la distribution de X_j dans la classe θ_k, dans un contexte opérationnel donné.

Théorème de Bayes généralisé

Mise en œuvre : fusion

- Capteurs indépendants :

$$m^{\Theta}[x_1, \ldots, x_J] = \bigcirc\!\!\!\!\cap_{j=1}^{J} m^{\Theta}[x_j] = \bigcirc\!\!\!\!\cap_{k=1}^{K} \overline{\{\theta_k\}}^{\prod_{j=1}^{J} pl_k(x_j)}.$$

- Capteurs dépendants :

$$m^{\Theta}[x_1, \ldots, x_J] = \bigcirc\!\!\!\!\wedge_{j=1}^{J} m^{\Theta}[x_j] = \bigcirc\!\!\!\!\cap_{k=1}^{K} \overline{\{\theta_k\}}^{\bigwedge_{j=1}^{J} pl_k(x_j)}.$$

Plan de la présentation

1. Théorie des fonctions de croyance
   - Concepts fondamentaux
   - Fusion d'informations
   - Décomposition canonique et règle prudente
2. Principes de construction de fonctions de croyance
   - Affaiblissement
   - Théorème de Bayes généralisé
   - Fonction de croyance prédictive

Fonction de croyance prédictive
Définition

- Soit X une variable aléatoire (définie en fonction du résultat d'une expérience répétable), de loi $\mathbb{P}_X$ inconnue.
- On a observé un échantillon indépendant, identiquement distribué de X : $X_1, \ldots, X_n$.
- Problème : construire une fonction de croyance $bel^\Omega[X_1, \ldots, X_n]$ relative à une prochaine réalisation de X $\longrightarrow$ fonction de croyance prédictive.
- Exemple : Ayant tiré n_B boules blanches et $n - n_B$ boules noires au cours de n tirages avec remise, fonction de croyance sur la couleur de la prochaine boule tirée.

Fonction de croyance prédictive

Propriétés souhaitées

- Propriété 1 (principe de Hacking) :

$$\forall A \subset \Omega, \quad bel^{\Omega}[X_1, \ldots, X_n](A) \xrightarrow{P} \mathbb{P}_X(A), \text{ as } n \to \infty,$$

- Propriété 2 (principe d'engagement minimal) :

$$\mathbb{P}\left(bel^{\Omega}[X_1, \ldots, X_n](A) \leq \mathbb{P}_X(A), \forall A \subset \Omega\right) \geq 1 - \alpha.$$

Fonction de croyance prédictive

Solutions

- X v.a. discrète, $\Omega = \{\xi_1, \ldots, \xi_K\}$: solution obtenue à partir d'une région de confiance sur les probabilités $p_k = \mathbb{P}(X = \xi_k)$:

$$\mathbb{P}\left(P_k^- \leq p_k \leq P_k^+, k = 1, \ldots, K\right) = 1 - \alpha$$

(T. Denoeux. International Journal of Approximate Reasoning, 2006).

- X v.a. continue, $\Omega = \mathbb{R}$: solution obtenue à partir d'une bande de confiance sur la fonction de répartition de X :

$$\mathbb{P}(F_*(x) \leq F_X(x) \leq F^*(x), \forall x \in \mathbb{R}) = 1 - \alpha.$$

(A. Aregui et T. Denoeux. Proceedings of ISIPTA '07, 2007).

Fonction de croyance prédictive

Exemple

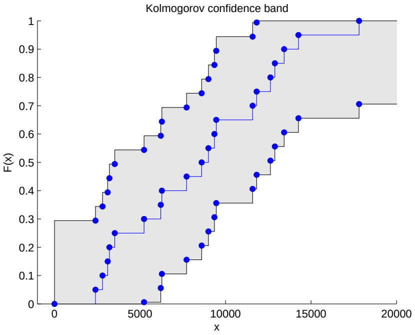
- Durées de vie de 20 roulement :

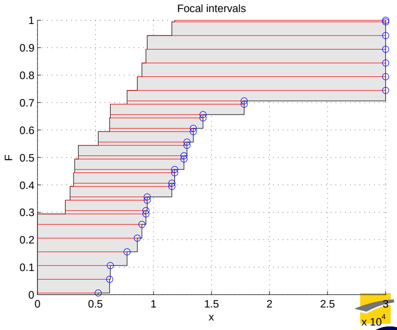
  2398, 2812, 3113, 3212, 3523, 5236, 6215,
  6278, 7725, 8604, 9003, 9350, 9460, 11584,
  11825, 12628, 12888, 13431, 14266, 17809.

- $X =$ durée de vie d'un roulement. Fonction de croyance sur X ?

Fonction de croyance prédictive

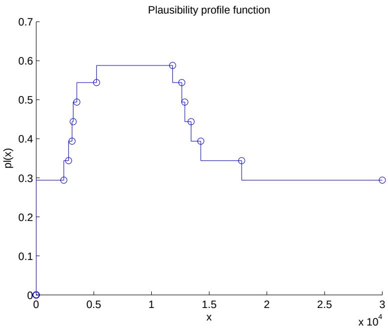
Bande de confiance de Kolmogorov

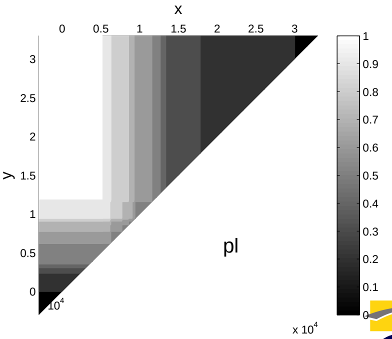




Fonction de croyance prédictive

Fonction de plausibilité





Conclusion (1/2)

- La théorie des fonctions de croyance (TFC) est un formalisme général qui étend les notions d'ensemble et de distribution de probabilité : représentation simultanée de l'imprécision et de l'incertitude.
- La TFC n'est pas un ensemble de recettes ad hoc : elle peut être construite rigoureusement à partir d'un petit nombre de principes (engagement minimal, Hacking).
- La TFC comporte un grand nombre (une infinité) de règles de fusion reposant sur différentes hypothèses relatives à la fiabilité et à l'indépendance des sources :
  - $\textcircled{\cap}$, $\textcircled{\cup}$
  - règle prudente $\textcircled{\wedge}$ pour la combinaison d'information non indépendantes.

Conclusion (2/2)

- L'affaiblissemnt contextuel permet de représenter dans un même cadre :
  - les informations fournies par des sources ;
  - des métaconnaissances sur la fiabilité (pertinence) de ces sources.
- Existence de méthodes générales et systématiques permettant de construire des fonctions de croyance à partir d'informations de natures différentes :
  - similarités,
  - vraisemblances,
  - échantillons statistiques,
  - etc.

Références I

cf. `http ://www.hds.utc.fr/~tdenoeux`

T. Denœux.
Conjunctive and Disjunctive Combination of Belief Functions Induced by Non Distinct Bodies of Evidence.
Artificial Intelligence, Vol. 172, pages 234-264, 2008.

D. Mercier, B. Quost et T. Denœux.
Refined modeling of sensor reliability in the belief function framework using contextual discounting.
Information Fusion, Vol. 9, pages 246-258, 2008.

Références II

cf. `http ://www.hds.utc.fr/˜tdenoeux`

T. Denœux.
Constructing Belief Functions from Sample Data Using Multinomial Confidence Regions.
International Journal of Approximate Reasoning, Vol. 42, Issue 3, Pages 228-252, 2006.

T. Denœux and P. Smets.
Classification using Belief Functions : the Relationship between the Case-based and Model-based Approaches.
IEEE Transactions on Systems, Man and Cybernetics B, Vol. 36, Issue 6, Pages 1395-1406, 2006.