Raisonnements approximatifs et systèmes complexes

Théorie des fonctions de croyance

Thierry Denœux

Université de Technologie de Compiègne
HEUDIASYC (UMR CNRS 6599)

Ingénierie Grands Projets et Systèmes Complexes
28 juin 2010

Plan de la présentation

Complexité et incertitude

- Les systèmes complexes sont par définition mal connus : lien entre les notions de complexité et d'incertitude.
- Les informations disponibles sur un système sont typiquement issues de deux types de sources :
  1. Des capteurs, soumis à des incertitudes de mesure et à des défaillances ;
  2. Des opinions d'experts.
- Ces informations sont typiquement hétérogènes, conflictuelles, imprécises, peu fiables, etc.
- Enjeu : modéliser et propager ces incertitudes dans les calculs et les raisonnements de manière à caractériser la confiance dans les résultats.

Formalismes classiques

- Deux formalismes classiquement utilisés en ingénierie :
  1. Théorie des probabilités :
     - Modèles stochastiques ;
     - Réseaux bayésiens, etc.
  2. Représensation ensembliste des incertitudes :
     - Calcul d'intervalles ;
     - Méthodes de statisfaction de contraintes, etc.

Théorie des probabilités

- Les probabilités sont typiquement utilisées pour représenter deux types d'incertitudes :
  - Les incertitudes aléatoires, dues à la variabilité de phénomènes répétables ou de caractéristiques au sein de populations ; probabilités objectives = fréquences ;
  - Les incertitudes épistémiques, dues au caractère incomplet de la connaissance ; probabilités subjectives, interprétation en termes de comportement en situation de pari.
- La première utilisation ne pose pas de problème, contrairement à la seconde.

Critiques de l'approche bayésienne

1. Incapacité à représenter l'ignorance :
   - Une distribution uniforme sur X induit généralement une distribution non uniforme sur $1/X$, X^2, $\log(X)$, etc. (paradoxe de Bertrand)
   - La théorie des probabilités n'est donc pas adaptée à la représentation d'états de connaissance dans lesquels on dispose de peu d'informations.
2. Arguments expérimentaux : en l'absence d'information, le comportement de décideurs ne peut pas toujours s'expliquer en supposant l'existence d'une distribution de probabilité subjective (paradoxe d'Ellsberg).

Représentation ensembliste

- La connaissance sur chaque variable X décrivant un système est représentée par un ensemble de valeurs possibles (par exemple : un intervalle, un pavé, une ellipse, etc.).
- Exemple : estimation à erreurs bornées
  - Soit un système dynamique décrit par les équations :

$$
\begin{aligned}
\mathbf{x}_{k+1} &= f(\mathbf{x}_k, \mathbf{u}_k) \\
\mathbf{y}_k &= g(\mathbf{x}_k, \mathbf{v}_k)
\end{aligned}
$$

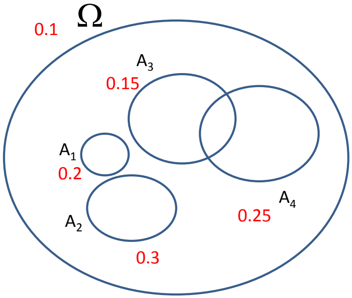
  - problème : étant donné des bornes sur les bruits et sur l'état initial, trouver un domaine garanti $\mathbf{X}_{k+1} \ni \mathbf{x}_{k+1}$.
- Avantage : simplicité des calculs dans certains cas (analyse par intervalles).
- Inconvénient : approche excessivement conservative.

Nouveaux formalismes

- Aucun des deux formalismes classiques n'est assez général.
- Depuis les années 1970, de nouveaux formalismes ont été proposés en Intelligence Artificielle et dans d'autres disciplines (Statistique, Economie) :
  - Théorie des possibilités ;
  - Théorie des probabilités imprécises ;
  - Théorie des fonctions de croyance (Dempster-Shafer).
- La théorie des fonctions de croyance, introduite par Dempster (1968) et Shafer (1976) et développée par Smets (1990) est l'un de formalismes les plus généraux, et le plus utilisé en ingénierie.
- Idée générale : synthèse entre l'approche probabiliste et l'approche ensembliste.

Plan de la présentation

Fonction de masse
Définition



- Ω : cadre de discernement.
- $m : 2^\Omega \rightarrow [0, 1]$ telle que

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

et $m(\emptyset) = 0$.
- Ensemble focal : A tel que $m(A) > 0$.
- $m(A)$ quantifie la croyance allouée à la proposition « $X \in A$ » et à aucune autre proposition plus précise.

Fonction de masse
Exemple

- Un meurtre a été commis. Il y a trois suspects : $\Omega = \{Pierre, Paul, Marie\}$.
- Un témoin a vu le meurtrier s'enfuir et prétend que c'est un homme. On sait que ce témoin est ivre 20 % du temps.
- L'information apportée par ce témoignage peut être représentée par

$$m(\{Pierre, Paul\}) = 0.8,$$

$$m(\Omega) = 0.2$$

- La masse de 0.2 n'est pas affectée à $\{Marie\}$, car le témoignage n'accuse absolument pas Marie !

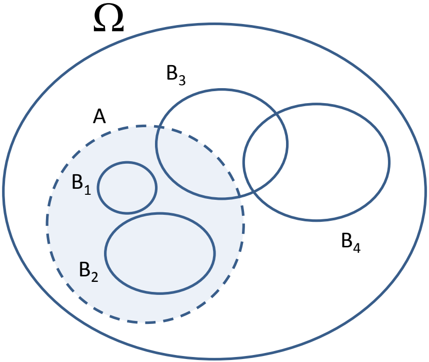
Fonction de masse
Cas particuliers

1. Un seul ensemble focal A tel que $m(A) = 1$:
   - Fonction de masse catégorique ;
   - m équivalente à un ensemble.
2. Ensembles focaux réduits à un seul élément :
   - Fonction de masse bayésienne ;
   - m est équivalent à une distribution de probabilité

Une fonction de masse est donc à la fois un ensemble généralisé et une distribution de probabilité généralisée

Fonctions de croyance et de plausibilité

Définitions



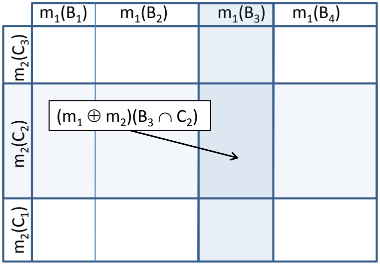
$$bel(A) = \sum_{B \subseteq A} m(B)$$

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B).$$

Fonctions de croyance et de plausibilité
Interprétation

- $bel(A)$: degré avec lequel les informations disponibles accréditent l'hypothèse « $X \in A$ ».
- $pl(A)$: degré avec lequel les informations disponibles ne discréditent pas l'hypothèse « $X \in A$ ».
- Intervalle $[bel(A), Pl(A)]$.
- Fonction de masse bayésienne : $bel = pl$, mesure de probabilité.

Règle de Dempster
Définitions



- Soient m_1 et m_2 deux fonctions de masse issues de sources indépendantes.

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{\sum_{B \cap C \neq \emptyset} m_1(B) m_2(C)}$$

- $\oplus$ commutatif et associatif.

Plan de la présentation

Plan de la présentation

Classification

Exemple : diagnostic par RdF

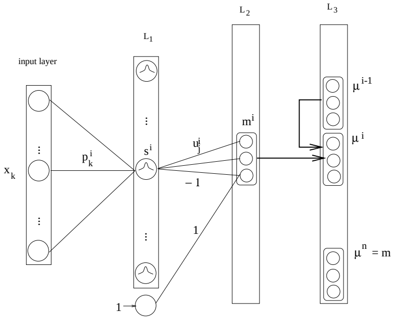
- Soit un système S caractérisé à chaque instant par une variable qualitative y décrivant son mode de fonctionnement (classe), et par un vecteur de mesures $\mathbf{x} \in \mathbb{R}^d$.
- Connaissance disponible : ensemble d'observations des variables $\mathbf{x}$ et y, réalisées sur le système à différents instants.
- Problème : construire un classifieur (règle prédisant la valeur de la variable y pour toute nouvelle valeur de la variable $\mathbf{x}$).

Classification

Intérêt des fonctions de croyance

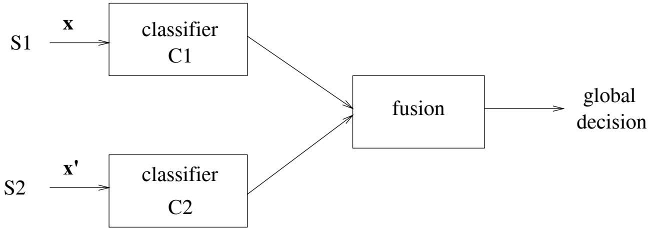
- Problèmes pour lesquels l'information disponible est relativement "pauvre" :
  - Ensemble d'apprentissage non exhaustif ;
  - Données d'apprentissage non totalement représentatives des données acquises en conditions opérationnelles ;
  - Données partiellement étiquetées (connaissance imparfaite des classes pour les données d'apprentissage), etc.
- Fusion d'informations issues de différentes sources (capteurs, experts, algorithmes d'apprentissage, etc.).

Réseau de neurone évidentiel



- $\Omega = \{\text{classes}\}$
- Apprentissage de prototypes $\mathbf{p}_i$ ayant des degrés d'appartenance aux classes.
- La similarité entre l'entrée $\mathbf{x}$ et un prototype $\mathbf{p}_i$ est considérée comme un élément d'évidence et induit une fonction de masse m_i.
- Combinaison des m_i par la règle de Dempster.

Problème



- $K = 2$ classes
- $\mathbf{x} \in \mathbb{R}^5$, $\mathbf{x}' \in \mathbb{R}^3$, distributions normales, independantes conditionnellement à la classe
- Ensemble d'apprentissage : $n = 60$, validation croisée : $n_{cv} = 100$
- test : 5000 exemples.

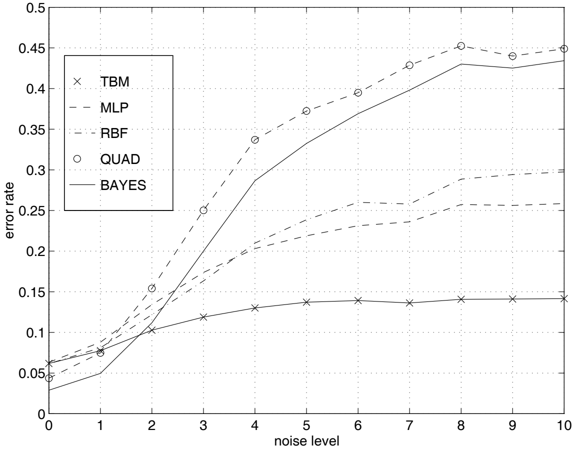
Résultats

Erreurs de tests : données non corrompues

Method	$\mathbf{x}$ seul	$\mathbf{x}'$ seul	$\mathbf{x}$ et $\mathbf{x}'$
RN év.	0.106	0.148	0.061
MLP	0.113	0.142	0.063
RBF	0.133	0.159	0.083
QUAD	0.101	0.141	0.049
BAYES	0.071	0.121	0.028

Résultats

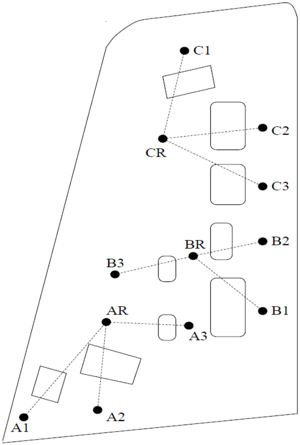
Erreurs de test : $\mathbf{x} + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$



Plan de la présentation

Surveillance acoustique de structure

Mechanical Systems and Signal Processing 23(6) :1792-1804, 2009

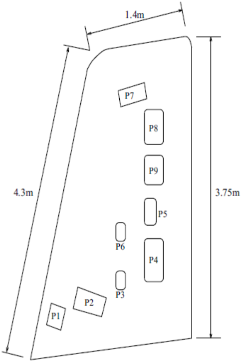


- 9 défauts (retrait des 9 panneaux d'inspection) $\rightarrow$ $K = 9$ classes.
- Mesures de transmissibilité par 12 capteurs acoustiques.
- 100 mesures pour chaque retrait d'un panneau d'inspection, avec 2 répétitions $\rightarrow$ 1800 observations.
- Sélection de caractéristiques (algorithme génétique) : 4 variables $\mathbf{x} \in \mathbb{R}^4$.

Résultats

Réseau de neurones évidentiel

29 prototypes, taux de bonnes classifications 89.7 %.

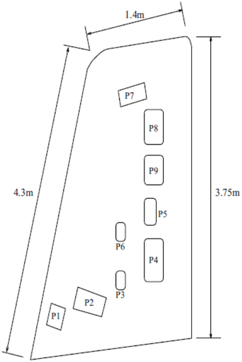


Prediction	1	2	3	4	5	6	7	8	9	Θ
True Class 1	54	5	5	0	0	0	2	0	0	0
True Class 2	0	63	0	0	2	0	0	0	0	1
True Class 3	6	1	56	2	0	0	0	0	0	1
True Class 4	5	0	1	55	0	3	0	2	0	0
True Class 5	0	0	0	0	65	0	0	1	0	0
True Class 6	2	2	2	4	0	54	1	0	0	1
True Class 7	0	1	1	0	0	0	61	2	1	0
True Class 8	0	0	1	0	1	0	0	62	1	1
True Class 9	0	0	0	0	0	0	0	3	63	0

Résultats

Réseau de neurones évidentiel + Perceptron multi-couches

Taux de bonnes classifications 92.3 %.



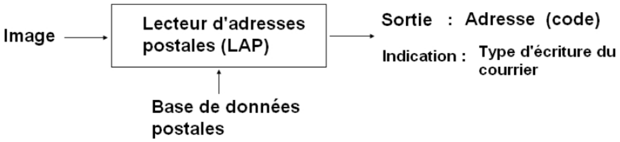
Prediction	1	2	3	4	5	6	7	8	9
True Class 1	63	2	0	0	0	0	1	0	0
True Class 2	0	64	0	0	2	0	0	0	0
True Class 3	0	1	55	6	0	0	2	2	0
True Class 4	0	1	2	59	0	1	0	3	0
True Class 5	0	0	0	0	65	0	0	0	1
True Class 6	1	2	0	6	0	57	0	0	0
True Class 7	0	0	0	0	0	0	60	5	1
True Class 8	1	0	0	0	1	0	0	63	2
True Class 9	0	0	0	0	0	0	0	4	62

Plan de la présentation

Fusion de lecteurs d'adresses postales

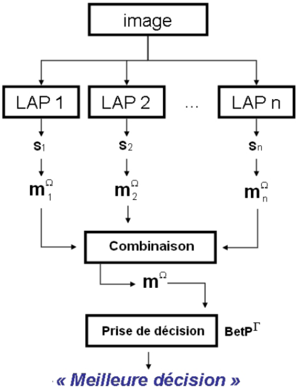
Expert Systems with Applications, 36(3) :5643-5653, 2009.

- Projet en collaboration avec la société Solystic.
- Problématique de la reconnaissance d'adresse postale : convertir automatiquement une image en adresse.



- Enjeu : maximiser le taux de reconnaissance tout en minimisant le taux de rejet.

Fusion de lecteurs d'adresses postales
Approche

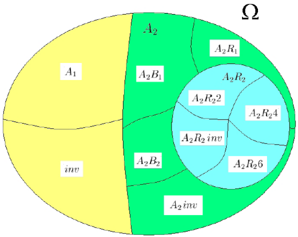


- Approche : fusionner les sorties de plusieurs LAP.
- Etapes de la modélisation :
  1. Définition du cadre de discernement
  2. Construction des masses
  3. Combinaison des masses
  4. Décision

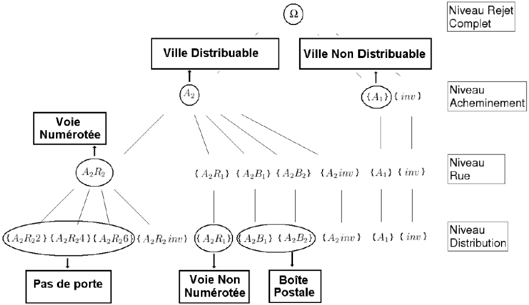
Fusion de lecteurs d'adresses postales

Définition du cadre de discernement

Ω = ensemble des adresses valides et invalides



Sorties d'un LAP :



Fusion de lecteurs d'adresses postales

Construction des masses, fusion et décision

- Construction des masses : utilisation d'information statistiques (matrices de confusion) à différents niveaux de granularité :
  - Acheminement ;
  - Rue ;
  - Distribution.
- Combinaison : règle de Dempster
- Décision :
  - Définition d'un cadre de pari
  - Prise en compte des coûts d'erreur et de rejet (total ou partiel).

Application de la règle de Dempster

Exemple : trois LAP_i ayant fourni les sorties s_i suivantes :

$s_1 = \{A_2 R_1 17\}$	$s_2 = \{A_1\}$	$s_3 = A_2$
$\begin{cases} m_1(\{A_2R_117\}) = 0.88 \\ m_1(A_2R_1) = 0.07 \\ m_1(A_2) = 0.03 \\ m_1(\Omega) = 0.02 \end{cases}$	$\begin{cases} m_2(\{A_1\}) = 0.9 \\ m_2(\Omega) = 0.1 \end{cases}$	$\begin{cases} m_3(A_2) = 0.95 \\ m_3(\{A_2 inv\}) = 0.04 \\ m_3(\Omega) = 0.01 \end{cases}$

Le calcul de $m = m_1 \oplus m_2 \oplus m_3$ donne (avant normalisation) :

$$\begin{array}{lcllcl} m(\{A_2R_117\}) & = & 0.0845 & m(\{A_1\}) & = & 0.0002 \\ m(A_2R_1) & = & 0.0067 & m(\Omega) & = & 0.0000 \\ m(\{A_2 inv\}) & = & 0.0002 & m(\emptyset) & = & 0.9036\,. \\ m(A_2) & = & 0.0048 & & & \end{array}$$

La masse importante allouée à l'ensemble vide reflète le conflit important entre les décisions fournies.

Application de la règle de Dempster

Exemple : trois LAP_i ayant fourni les sorties s_i suivantes :

$s_1 = \{A_2 R_1 17\}$	$s_2 = \{A_1\}$	$s_3 = A_2$
$\begin{cases} m_1(\{A_2 R_1 17\}) = 0.88 \\ m_1(A_2 R_1) = 0.07 \\ m_1(A_2) = 0.03 \\ m_1(\Omega) = 0.02 \end{cases}$	$\begin{cases} m_2(\{A_1\}) = 0.9 \\ m_2(\Omega) = 0.1 \end{cases}$	$\begin{cases} m_3(A_2) = 0.95 \\ m_3(\{A_2 inv\}) = 0.04 \\ m_3(\Omega) = 0.01 \end{cases}$

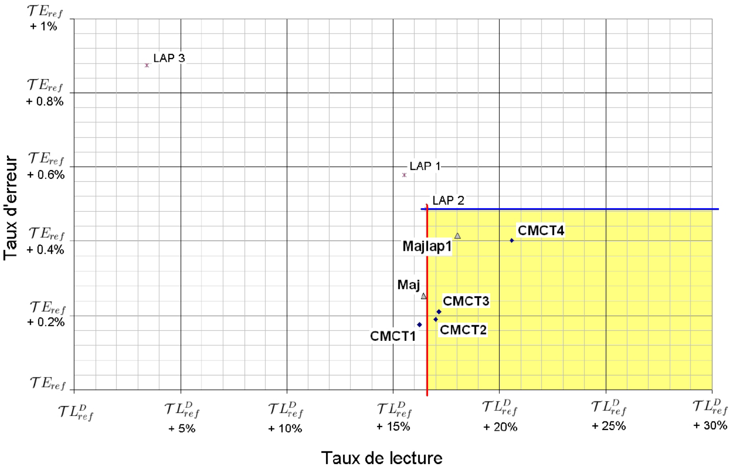
Le calcul de $m = m_1 \oplus m_2 \oplus m_3$ donne (avant normalisation) :

$$\begin{array}{lcllcl} m(\{A_2 R_1 17\}) & = & 0.0845 & m(\{A_1\}) & = & 0.0002 \\ m(A_2 R_1) & = & 0.0067 & m(\Omega) & = & 0.0000 \\ m(\{A_2 inv\}) & = & 0.0002 & m(\emptyset) & = & 0.9036\ . \\ m(A_2) & = & 0.0048 & & & \end{array}$$

La masse importante allouée à l'ensemble vide reflète le conflit important entre les décisions fournies.

Fusion de lecteurs d'adresses postales

Exemple de résultat



Conclusions (1/2)

- La théorie des fonctions de croyance généralise à la fois
  - la théorie des probabilité et
  - les approches basées sur une représentation ensembliste des informations.
- Il s'agit donc d'un langage très général permettant
  - de construire des modèles de systèmes complexes en présence de différents types d'incertitudes (aléatoires et épistémiques) ;
  - de fusionner des informations issues de sources imparfaites et hétérogènes.

Conclusions (2/2)

- Nombreuses applications développées à Heudiasyc :
  - Diagnostic de circuits de voie (SNCF, INRETS) ;
  - Prise en compte de la fiabilité et de la sincérité des sources en fusion d'informations (Thalès) ;
  - Gestion des risques en distribution d'eau potable (Suez Environnement) ;
  - Fouille de données, apprentissage (méthodes d'ensembles en classification et en clustering ; classification multi-label, etc.).
- Travaux théoriques :
  - Développement de nouveaux opérateurs de combinaison (fusion d'informations issues de sources non indépendantes) ;
  - Manipulation de fonctions de croyance dans de très grands espaces (partitions, relations d'ordre, etc.).
  - Méthodes formelles d'élicitation, etc.