# Votes et fonctions de croyance pour la fusion de décisions

Arnaud MARTIN
ENSIETA – $E^3I^2$ - EA3876
Arnaud.Martin@ensieta.fr

# Plan

- Qu'est-ce que la fusion ?
- Généralités
- Méthodes des votes
- Théorie des fonctions de croyance

# Plan

- **Qu'est-ce que la fusion ?**
- Généralités
- Méthodes des votes
- Théorie des fonctions de croyance

# Incertitude linguistique

Airbus A320 en x,y

Mauvaise météo

Bonne météo

Proposition incertaine

Proposition sûre, certaine

# Imprécision linguistique

Le trafic est dense au péage autoroutier

Proposition imprécise

# Imprécision linguistique

Dix voitures attendent au péage

Proposition précise

# Buts de la fusion ?

Obtenir une information aussi sûre et précise que possible dans toutes les conditions d'observation en profitant de la complémentarité et de la redondance d'informations élémentaires lesquelles ne sont ni complètement fiables ni totalement précises

# Objectifs de la fusion

Supposons que l'ensemble des classifieurs est totalement sûre et précis

C'est un A380

C'est un A380

C'est un A380

C'est un A380

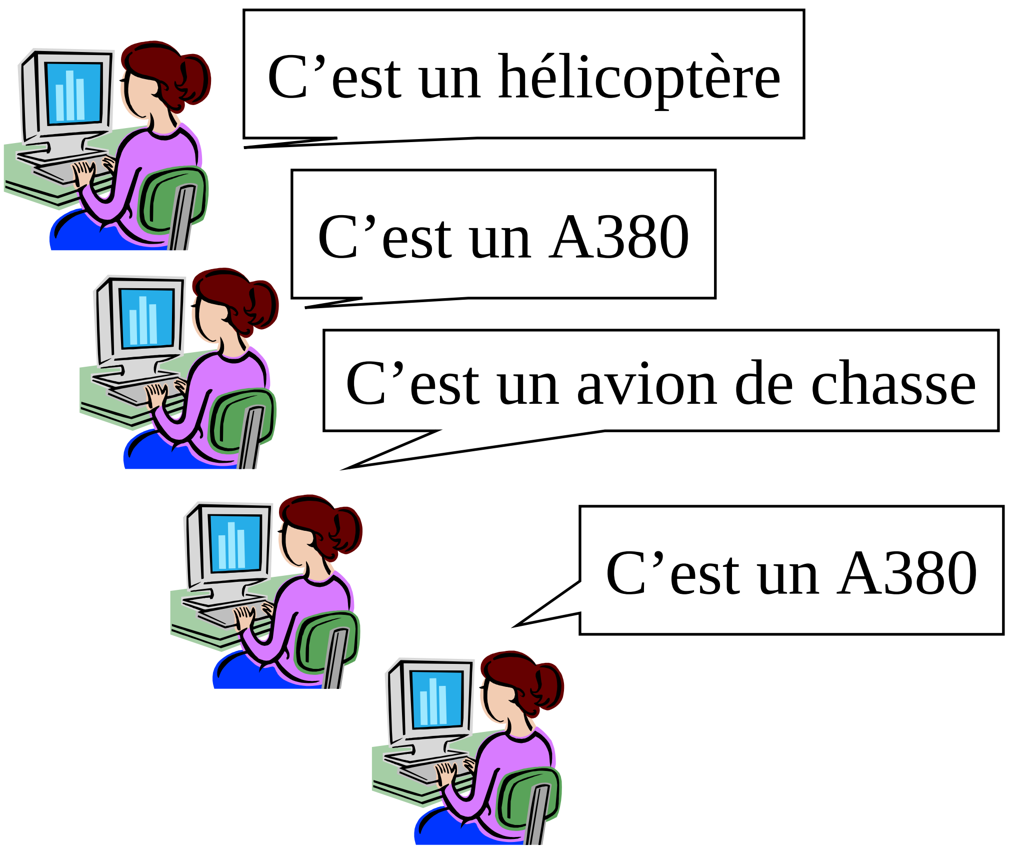
**FUSION**

# Objectifs de la fusion de données

Une seule source est suffisante

# Objectifs de la fusion de données

Si l'ensemble des sources n'est ni totalement sûr et précis





**FUSION**

**Nécessité de développer des techniques adaptées**

# Fiabilité des sources : cotation de l'information

A

Je suis sûr que c'est un avion de chasse

A se trompe souvent
B ne se trompe jamais

B

C'est peut être un A380

C'est peut être un A380

## Fusion de Classifieurs

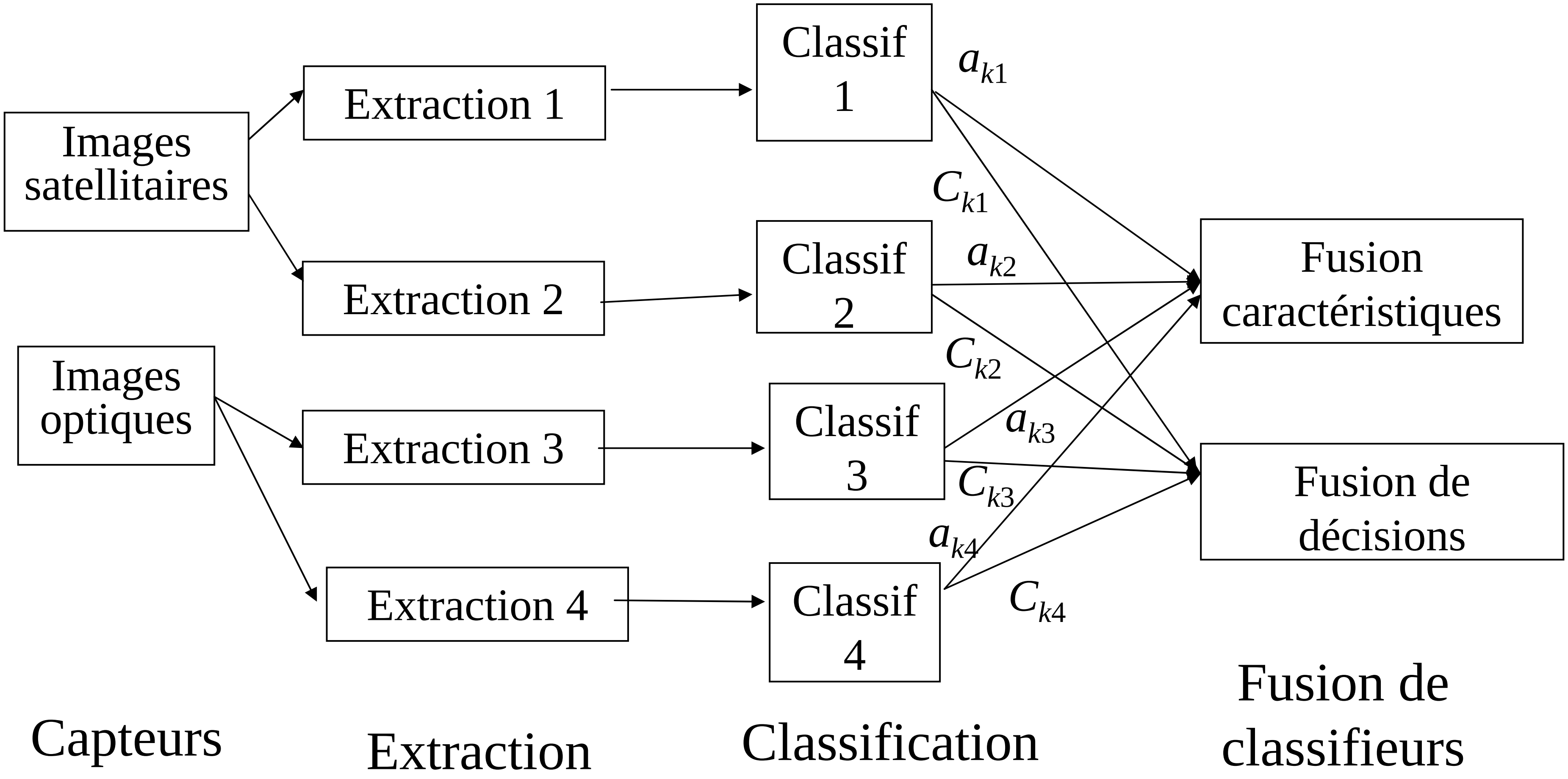
Soit $C_1,\ldots, C_m$, $m$ classifieurs qui caractérisent $n$ avions. Pour un même signal ou des signaux différents (optique, radar, ..) chaque classifieur décide du type d'avion.

But : obtenir une décision plus sûre et plus précise du type d'avion en considérant l'ensemble des décisions prises au niveau de chaque classifieur

Fusion de décisions
Fusion de caractéristiques

Exemple de fusion

# Fusion de Classifieurs



# Problèmes



• Comment combiner les informations dans les deux cas ?

• Comment prendre la décision ?

**Sous problème dépendant de l'application**

• Comment modéliser l'information ?

*i.e.* choix d'un formalisme

• Comment estimer les paramètres du modèle

# Plan

- Qu'est-ce que la fusion ?
- **Généralités**
- Méthodes des votes
- Théorie des fonctions de croyance
- Classification multistratégies

# Plan : Généralités

- Définitions de la fusion d'informations
- Domaines d'application
- Typologie de l'information
- Architecture de fusion
- Modèles de fusion

# Définitions de la fusion d'informations

Fusionner : regrouper, combiner des informations

Informations : tout élément pouvant être codé

Déf. 1 (groupe de travail FUSION) : *regrouper* des informations issues de *plusieurs sources* d'informations et exploiter l'information regroupée

Déf. 2 (GDR-ISIS) : *combiner* des informations issues de *plusieurs sources* afin d'améliorer la prise de *décision*

# Problèmes : les imperfections des données

– les supprimer

– les tolérer en robustifiant les algorithmes

– les modéliser

# Domaines d'application

- Fusion en traitement du signal
- Fusion en traitement de l'image
- Fusion en robotique

**Buts**

- Pour reconnaissance (classification, identification, détection)
- Estimation
- Association

# Typologie de l'information

- Type d'information
- Niveaux de fusion
- Les sources d'information
- Imperfections des données
- Informations supplémentaires

# Type d'information



- Numérique

  Sous forme de nombres (quantitatif et qualitatif)

  *ex* niveaux de gris, intensité du signal, temps d'arrivées…

- Symbolique

  Sous forme de symboles, propositions, taxonomie, règles d'associations

  *ex* règles d'associations (*si ça vole*, *c'est un avion*), propositions (*ce qui est grand n'est pas petit*), taxonomie (*le chat est un mammifère),* …

# Niveaux de fusion

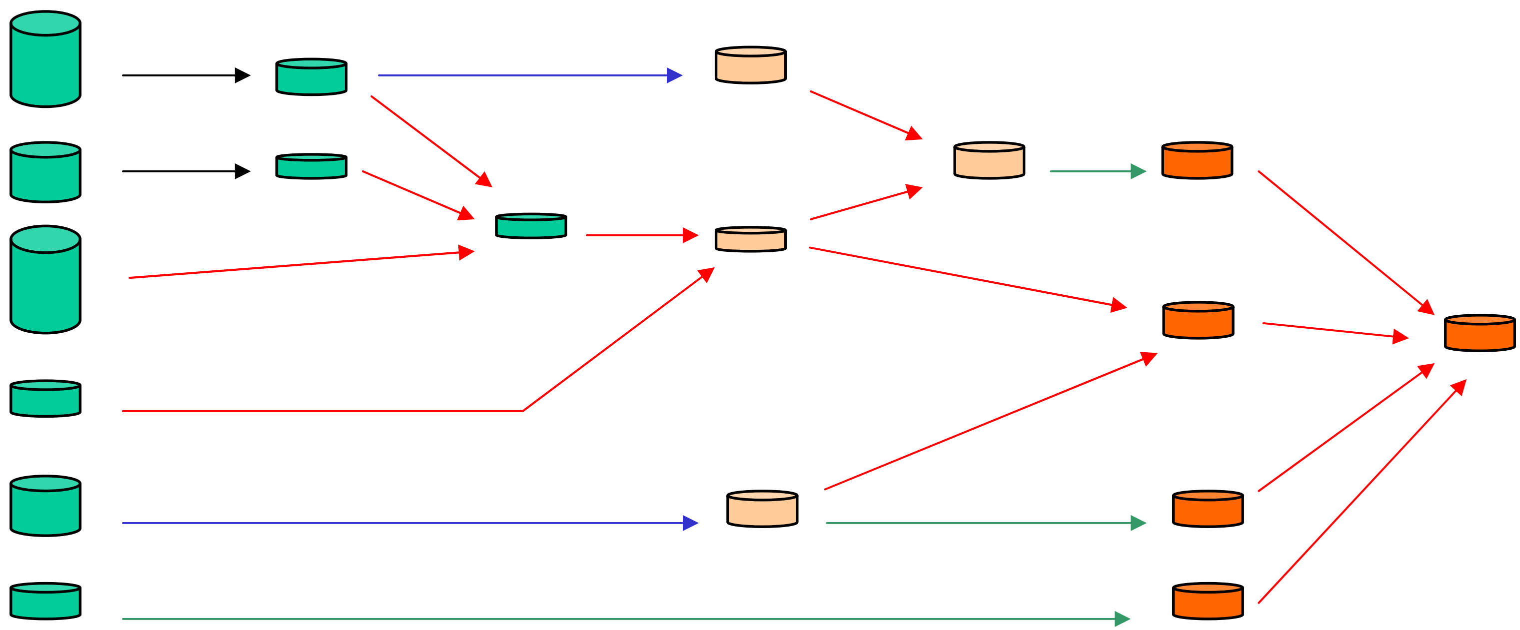


- Fusion de données
- Fusion de caractéristiques
- Fusion de décisions
- Fusion de modèles

# Niveaux de fusion

## Typologie 4/9

Différentes sources



Sélection, transformation (Analyse de données)

Données (signal, informations, etc..)

Extraction d'informations

Caractéristiques (images, formes, etc..)

Classification, Détection, Identification, Reconnaissance

Décisions (numériques, symboliques)

FUSION de données, de caractéristiques, de décisions

# Sources d’information



- Indépendances (hypothèse souvent faite)
  - Indépendance statistique
  - Indépendance cognitive
- Fiabilité (souvent difficile à quantifiée)
  - Fiabilité de la source à donner une information sûre
  - Fiabilité de la source à donner une information
- Informations influant sur le choix de la modélisation des informations à fusionner

# Imperfections des données



- Incertitude

  Caractérise le degré de conformité à la réalité. Ex : *cette lettre arrivera demain*

- Imprécision

  Caractérise le contenu de l'information (défaut quantitatif de connaissance). Ex : *cet homme est grand*

  Incertaine et imprécise : *il pleuvra beaucoup demain*

- Incomplétude

  Manque d'informations apporté par la source. Ex : données manquantes dues à la transmissions

# Imperfections des données



- Redondance

  Les sources apportent plusieurs fois les mêmes informations. Ex : observations vidéo d'un même objet

- Complémentarité

  Les sources apportent des informations sur des grandeurs différentes (caractéristiques différentes). Ex :

  – informations I.R et informations visibles

  – informations GPS et informations visibles

- Conflit

  Les sources fournissent des informations en contradiction

  Difficile à détecter

# Imperfections des données

Typologie 8/9

- Ambiguïté

  Une information conduit à 2 interprétations (peut provenir d'une autre imperfection : incertitude, imprécision, conflit). Ex : forme d'un avion trop proche d'un autre ne permet pas de l'identifier

# Informations supplémentaires

Typologie 9/9

- Sur les informations à combiner :

  sur les sources : sur leur dépendance, leur fiabilité,
  des préférences sur les informations à combiner

- Sur le domaine

- Pas toujours dans le même formalisme que les informations à combiner
- Peuvent aider dans les différentes étapes de la fusion

# Architecture de fusion

## Position du problème

$m$ sources $S_1, S_2, \ldots, S_m$ qui doivent prendre une décision sur une observation $x$ dans un ensemble de $n$ classes

$x \in C_1, C_2, \ldots,$ ou $C_n$ classes

$$
\begin{array}{cc}
 & \begin{array}{ccccc} C_1 & \cdots & C_i & \cdots & C_n \end{array} \\
\begin{array}{c} S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{array} &
\begin{bmatrix}
M_1^1(x) & \cdots & M_i^1(x) & \cdots & M_n^1(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^j(x) & \cdots & M_i^j(x) & \cdots & M_n^j(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^m(x) & \cdots & M_i^m(x) & \cdots & M_n^m(x)
\end{bmatrix}
\end{array}
$$

# Architecture de fusion 2/9

$$
\begin{array}{c c}
 & \begin{array}{ccccc} C_1 & \ldots & C_i & \ldots & C_n \end{array} \\
\begin{array}{c} S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{array} &
\begin{bmatrix}
M_1^1(x) & \ldots & M_i^1(x) & \ldots & M_n^1(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^j(x) & \ldots & M_i^j(x) & \ldots & M_n^j(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^m(x) & \ldots & M_i^m(x) & \ldots & M_n^m(x)
\end{bmatrix}
\end{array}
$$

$\xrightarrow{\text{Fusion}}$ La classe

# Architecture de fusion 3/9

$$\begin{array}{ccc}
 & \begin{matrix} C_1 & \cdots & C_i & \cdots & C_n \end{matrix} & \\
\begin{matrix} S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{matrix} &
\begin{bmatrix}
M_1^1(x) & \cdots & M_i^1(x) & \cdots & M_n^1(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^j(x) & \cdots & M_i^j(x) & \cdots & M_n^j(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^m(x) & \cdots & M_i^m(x) & \cdots & M_n^m(x)
\end{bmatrix} &
\begin{matrix} C_4 \\ \vdots \\ C_1 \\ \vdots \\ C_6 \end{matrix}
\end{array}
\xrightarrow{\text{Fusion}} \text{La classe}$$

# Architecture de fusion 4/9

$$
\begin{array}{cc}
 & \begin{array}{ccccc} C_1 & \cdots & C_i & \cdots & C_n \end{array} \\
\begin{array}{c} S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{array} &
\begin{bmatrix}
M_1^1(x) & \cdots & M_i^1(x) & \cdots & M_n^1(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^j(x) & \cdots & M_i^j(x) & \cdots & M_n^j(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
M_1^m(x) & \cdots & M_i^m(x) & \cdots & M_n^m(x)
\end{bmatrix} \\
 & \begin{array}{ccccc} M_1 & \cdots & M_i & \cdots & M_n \end{array}
\end{array}
$$

$$\Big\downarrow \ \text{Fusion}$$

La classe

# Étapes de la fusion

1. Modélisation
2. Estimation
3. Combinaison
4. Décision

## Architecture de fusion 6/9

- Étape déterminante
- Choix du formalisme (représentation des informations à fusionner)
- Informations supplémentaires peuvent guider ce choix
- **Exemple**

  Chaque source $S_j$ fournit une information représentée par $M_i^j$ sur la classe $C_i$

  $M_i^j$ peut être une distribution, une formule, fonction de coût, etc…

# Estimation

- Phase souvent nécessaire dépend de la modélisation (du formalisme)
- Par exemple pour l'estimation des distributions, ou de la fiabilité des sources

- Informations supplémentaires peuvent aider

• Phase de regroupement des informations

• Choix d'un opérateur adapté au formalisme de la modélisation

• Donne un résultat de même nature que les informations combinées

• Propriétés intéressantes

associativité, commutativité, idempotence, adaptativité

• Propriétés limitatives

additivité

• Informations supplémentaires peuvent aider

# Décision

- Dernière phase de la fusion
- Passage des informations fournies par les sources à la décision finale
- Choix du critère en fonction du formalisme
- Minimisation ou maximisation d'une fonction issue de la combinaison

- Fournit à l'expert la *meilleure* décision (avec un indice de qualité)

# Les modèles de fusion

**Représentations numériques**

– Méthodes de votes

fonctions indicatrices

– Approche bayésienne (probabiliste et statistique)

probabilités d'évènements (fréquences)

– Théorie des ensembles floues et des possibilités

nécessité et possibilité

– Théorie des croyances (Dempster-Shafer)

plausibilité et croyance

# Plan

- Qu'est-ce que la fusion ?
- Généralités
- **Méthodes des votes**
- Théorie des fonctions de croyance

# Méthodes de vote

Approche simple et intuitive

**Caractéristiques de cette approche**

• Peut s'appliquer sans *a priori*

• Ajout de fiabilité possible (moyennant apprentissage)

• Plus adaptée pour la fusion haut niveau (de décisions)

Mettons nous à la place d'un juge d'instruction $E$ qui instruit un crime récent. Bien que dépourvu de tout indice matériel, le juge a cependant la chance de disposer de $n$ témoins oculaires indépendants ($T_1$, $T_2$, …, $T_n$). L'accusé $x$ peut être coupable, innocent ou complice des faits qui lui sont reprochés. Nous noterons ces trois classes respectivement par $C_1$, $C_2$ et $C_3$. Dans le cas où le juge ne peut prendre de décision, nous noterons que $x$ appartient à la classe $C_4$.

- Supposons que $n$=15, 5 témoins accusent $x$ d'être coupable, 6 disent qu'il est innocent, et 4 qu'il est complice.
- Si 6 disent qu'il est coupable, 6 disent qu'il est innocent, et 3 qu'il est complice.

Parmi les 15 témoins oculaires, un témoin est policier. Les policiers étant assermentés, le juge a une confiance absolue en son témoignage. Les 14 autres témoins sont des voyous, tous aux casiers judiciaires bien remplis. Ces 14 témoins fournissent à l'accusé un alibi, tandis que le policier accuse *x*. Attention un témoin peut ne pas être fiable, mais dire vrai (cas d'un témoin bandit qui dit la vérité).

Le juge va donc attribuer un *degré de fiabilité* différent à chaque témoin.

**Modélisation :** fonction indicatrice :

$$M_i^j(x) = \begin{cases} 1 & \text{si} \quad S_j : x \in C_i \\ 0 & \text{sinon} \end{cases}$$

**Estimation :** $\alpha_{jk}$ fiabilité d'une source pour une classe donnée

**Combinaison :** $$M_k^E(x) = \sum_{j=1}^{m} \alpha_{jk} M_i^j(x)$$

**Décision :**

$$x \in C_k \text{ si } M_k^E(x) = \max_i M_i^E(x) \geq c.m + b(x)$$

$$x \in C_{n+1} \text{ sinon i.e. non décision}$$

$$c \in [0,1],\ b(x) \text{ fonction de } M_k^E(x)$$

# Résultat

## Hypothèse :

indépendance statistique des sources

Même probabilité de réussite $p$

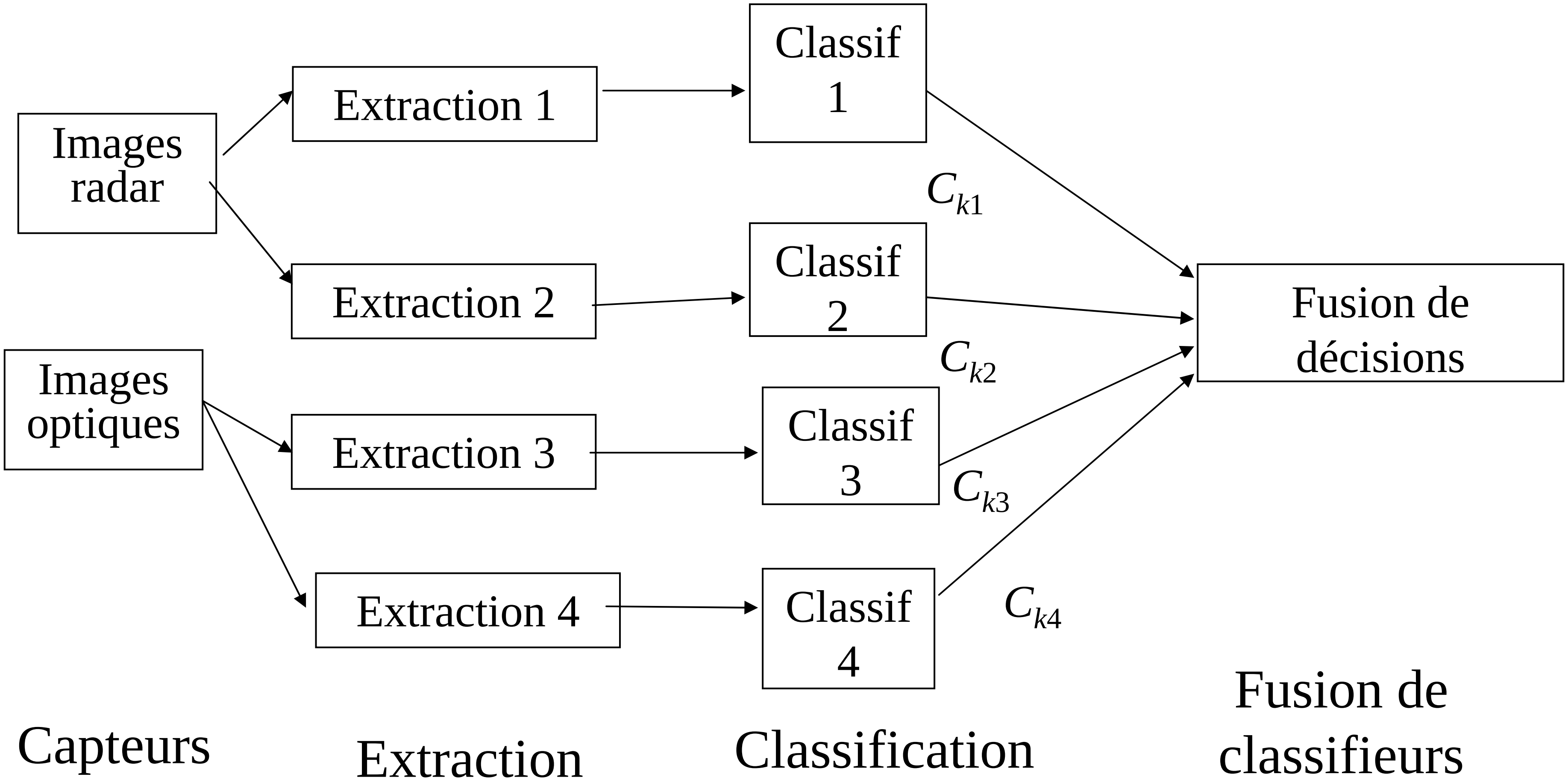
Et $m$ impair $>2$

## Alors :

- Si $p>0.5$, $P_E$ tend vers 1 avec $m$
- Si $p<0.5$, $P_E$ tend vers 0 avec $m$
- Si $p=0.5$, $P_E=0.5$ pour tout $m$

où $P_E$ est la proba d'une réponse juste après fusion

# Fusion de Classifieurs

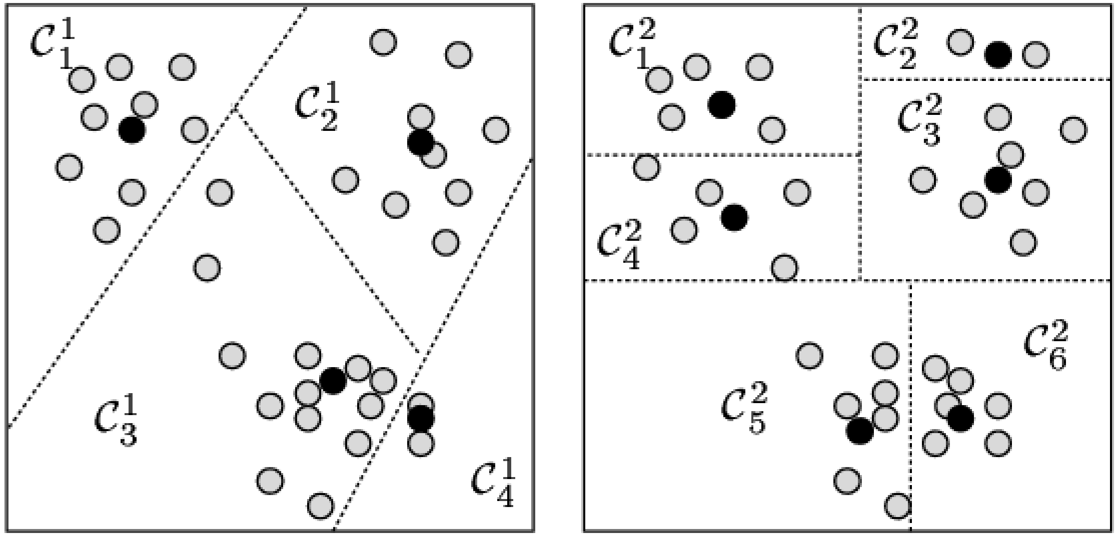


- $M_i^j(x)$ est la réponse du classifieur $j$ pour l'observation $x$ : 0 si $x$ n'appartient pas à la classe $i$, 1 sinon

- La fiabilité d'une source pour une classe donnée peut être estimée à partir des matrices de confusion de chaque classifieur

# Fusion de classifieurs
## Non supervisés

## Méthodes de vote




Gançarski et Wemmert, 2002

- Cas de la classification non supervisée : chaque classifieur s’exprime sur un cadre de discernement différent

- Nécessité de mettre en correspondance (association) les réponses des classifieurs avant de les combiner

# Fusion de classifieurs
Gancarski et Wemmert 2002

## Méthodes de vote

- Association : mesure de similarité entre les classifieurs $S_i$ et $S_j$ à partir des

$$\alpha_{k,l}^{i,j} = \frac{|C_k^i \cap C_l^j|}{|C_k^i|}$$ $S_i$ dit $C_k$ et $S_j$ dit $C_l$

- Similarité entre classes :

$$S(C_k^i, C_l^j) = \alpha_{k,l}^{i,j} \alpha_{l,k}^{j,i}$$

- Fonction de correspondance

$$F(C_k^i) = argmax_{C_l^j} S(C_k^i, C_l^j)$$

- Vote sur les classes raffinées correspondantes

# Fusion de classifieurs
Gancarski et Wemmert 2005

# Méthodes de vote

- Problème de conflit si $|C_k^i \cap C_l^j| = \emptyset$

pour beaucoup de classes

- Définition d'une mesure de concordance et qualité locale entre deux classifieurs avec l'introduction d'informations données par l'expert (somme pondérée de la mesure de similarité et d'un critère de qualité)

- Problème pour les classes empiétantes
- Le raffinement des classes se fait uniquement sur les classes existantes. Une classe d'un classifieur est associée à une unique autre classe raffinée
- Mesure de similarité symétrique (2002)

$$S(C_k^i, C_l^j) = S(C_l^j, C_k^i)$$

Si une classe *A* en contient parfaitement 2 autres *B* et *C*, *F*(*B*)=*A* et *F*(*C*)=*A*, n'entraînent pas *F*(*A*)=*B* et *F*(*A*)=*C*

# Plan

- Qu'est-ce que la fusion ?
- Généralités
- Méthodes des votes
- **Théorie des fonctions de croyance**

# Théorie des fonctions de croyance
# Théorie de Dempster-Shafer
# Théorie de l'évidence

# Plan

- Limites bayésiennes
- Principes
- Modélisation
- Estimation
- Combinaison
- Décision
- Exemple pour l'association de classes
- Réflexions

# Limites bayésiennes

- Représente mal l'imprécision (confusion de l'incertain et de l'imprécision)
- Modélisation : difficile de modéliser des connaissances qui ne se traduisent pas par des probabilités, ou l'absence de connaissances (*ex* : Sirius)
- Contraintes sur mesures (probabilités, additivité) et sur les classes (exhaustivité et exclusivité)

**Exemple de Smets : domaine diagnostique médical**

Si un symptôme *s* est toujours présent quand un patient a une pathologie *A* ($p(s/A)=1$), et que l'on observe ce symptôme *s*, alors la probabilité pour que le patient ait A augmente (car $p(A/s)=p(A)/p(s)$ donc $p(A/s)>=p(A)$).

La contrainte d'additivité impose alors que la probabilité pour que le patient n'ait pas *A* diminue :

$p(\text{non}A/s)=1-p(A/s)$ donc $p(\text{non}A/s)<=p(\text{non}A)$

Alors qu'il n'y a pas de raison, si le symptôme *s* peut aussi être observé dans d'autres pathologies.

# Conclusions sur approches bayésiennes

## Limites

• Optimale lorsque l'*a priori* est parfaitement connu (jamais le cas en pratique)

• Base mathématique solide

• Outils riches et rigoureux pour la modélisation et l'apprentissage (connaissance acquise par de nombreuses expériences)

• Combinaison permettant la mise à jour

• Hypothèse d'indépendance pas obligatoire

# Principe générale de la théorie des fonctions de croyance

- Principe repose sur la manipulation de fonctions sur des sous-ensembles (au lieu de singleton dans la théorie des probabilités) à valeur dans [0,1]
- Espace de discernement $\Theta=\{C_1, C_2, \ldots, C_n\}$ où $C_i$ représente une hypothèse
- On va définir des fonctions sur l'espace

$$2^{\Theta}=\{\emptyset, \{C_1\}, \{C_2\}, \ldots, \{C_n\}, \{C_1,C_2\}, \ldots, \Theta\}$$

ensemble de toutes les disjonctions possibles des hypothèses

## Extension de la DSmT à $D^{\ominus}$.

## Principe

Extension récente des fonctions sur l'espace

$$D^{\ominus} = \{\emptyset, \{C_1\}, \{C_2\}, ..., \{C_1 \cap C_2\}, ..., \{C_1 \cup C_2\}\}$$

ensemble de toutes les disjonctions et conjonctions possibles des hypothèses

Intérêt : Exemple d'une route inondée

$S_1$ dit *Rivière* et $S_2$ dit *Route*

Les 2 classifieurs n'ont pas tords

# Modélisation

Définition de fonctions qui permettent de représenter l'imprécision et l'incertitude

– Fonctions de masse $m$

– Fonction de crédibilité $Cr$

– Fonction de plausibilité $Pl$

– Fonction de communalité $Q$ et Fonction d'implicabilité $b$

masse élémentaire ou masse de croyance définie sur $2^{\Theta}$ à valeur dans [0,1] pour une source $S_j$

On impose en général

$$\sum_{A \subseteq D} m_j(A) = 1$$

$m_j(\varnothing)=0$ (hypothèse de monde clos, fermé)

$m_j(A)$ caractérise un degré de croyance en $A$

Un *élément focal* est un élément $A$ de $2^{\Theta}$ tel que $m_j(A)>0$. La réunion des éléments focaux est le *noyau*

## *Cas particulier*

- Si les seuls éléments focaux sont les $C_i$ alors $m$ est une probabilité
- $m_j(\Theta)=1$ : ignorance totale de $S_j$
- $m_j(A)=1$ : $S_j$ a une connaissance imprécise (elle ne croit qu'en $A$ : réunion de classes)
- $m_j(C_i)=1$ : $S_j$ a une connaissance précise (elle croit pleinement en $C_i$)
- $m_j(A)=s$ et $m_j(\Theta)=1-s$ : $S_j$ a une connaissance incertaine et imprécise (elle croit en partie en $A$ mais rien de plus)

Regroupe les masses des éléments inclus

Degré de croyance

Pour tout $A$ de $2^{\Theta}$

$$Cr_j(A) = \sum_{B \subseteq A, B \neq \emptyset} m_j(B)$$

et $Cr_j(\Theta)=1- m_j(\emptyset)$ (=1 en monde fermé)

Fonction croissante de $2^{\Theta}$ sur [0,1]

Cr(A1,A2,A3)>=Cr(A1)+Cr(A2)+Cr(A3)-Cr(A1 et A2) – Cr(A1 et A3) – Cr(A2 et A3)+ Cr(A1 et A2 et A3)

# Fonctions de crédibilité



- Ces fonctions mesurent toute la croyance sur $A$ *i.e.* l'intensité que les informations fournies par $S_j$ soutiennent la proposition $A$ (croyance minimum)
- Avoir $m_j(A)=0$ ne signifie pas que $A$ est impossible (si $Cr_j(A)>0$), mais qu'on ne sait pas affecter un degré précisément à $A$
- On a aussi pour tout $A$ de $2^\Theta$

$$m_j(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Cr_j(A)$$

# Fonctions de plausibilité



- Regroupe les masses des éléments intersectés
- Pour tout $A$ de $2^{\Theta}$

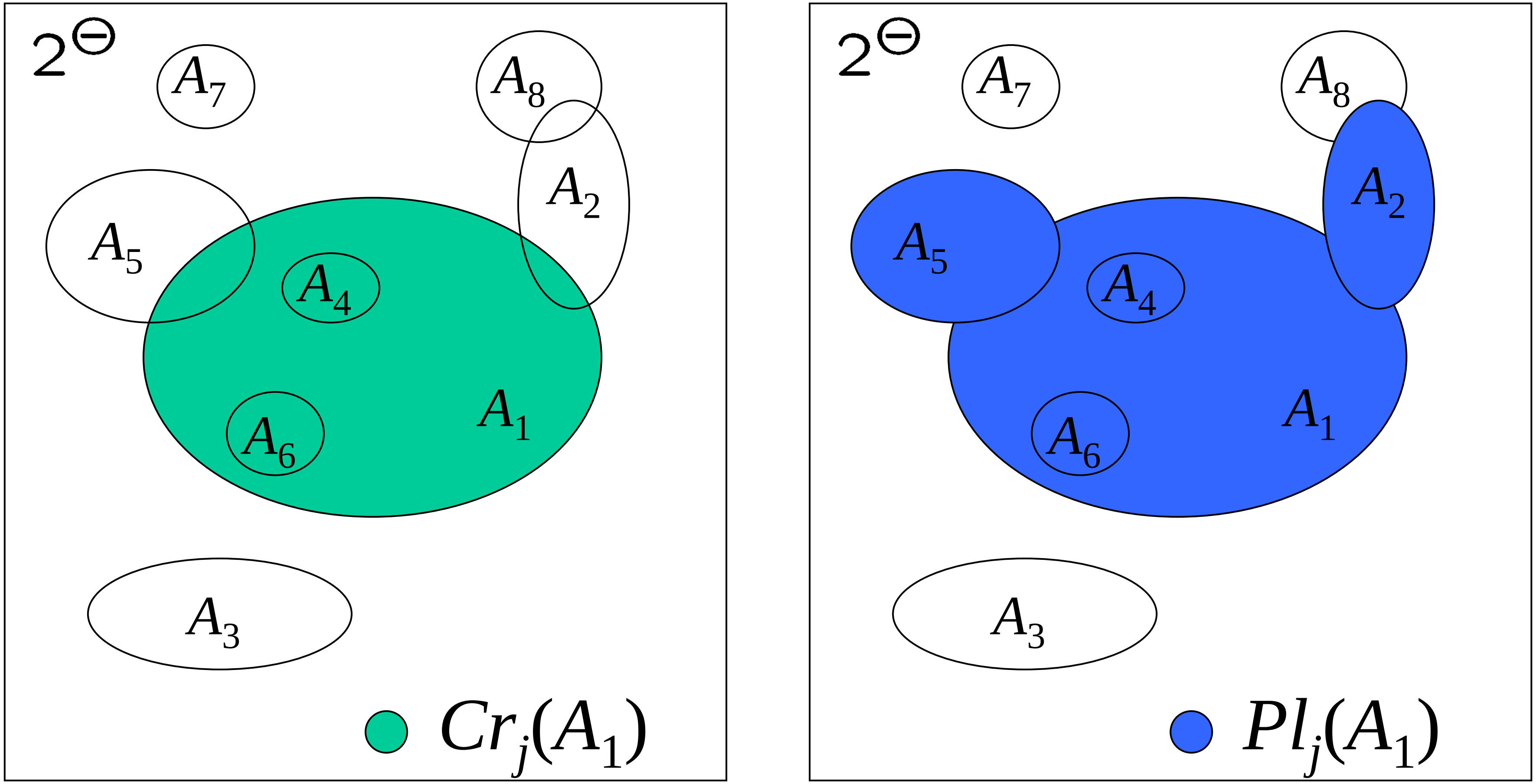
$$Pl_j(A) = \sum_{B \cap A \neq \emptyset} m_j(B) = Cr_j(\Theta) - Cr_j(A^c)$$

(=1-$Cr_j(A^c)$) en monde fermé)

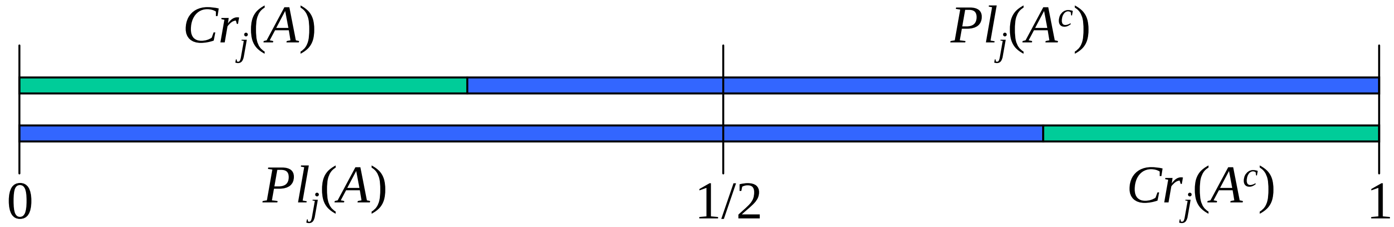
- Mesure l'intensité avec laquelle on trouve $A$ plausible, *i.e.* avec laquelle on ne doute pas de la proposition $A$. C'est aussi la croyance maximum en $A$.

# Propriétés





$2^\Theta$
$A_7$
$A_8$
$A_2$
$A_5$
$A_4$
$A_6$
$A_1$
$A_3$
$Cr_j(A_1)$
$2^\Theta$
$A_7$
$A_8$
$A_2$
$A_5$
$A_4$
$A_6$
$A_1$
$A_3$
$Pl_j(A_1)$

$Cr_j(A)$
$Pl_j(A^c)$
0
$Pl_j(A)$
1/2
$Cr_j(A^c)$
1

# Propriétés



Pour tout $A$ de $2^{\Theta}$ on a :

- $Pl_j(A) \geq Cr_j(A)$

en monde fermé :

- $Cr_j(A) + Cr_j(A^c) \leq 1$

- $Pl_j(A) + Pl_j(A^c) \geq 1$

- $Cr_j(A) + Cr_j(A^c) = 1 \Leftrightarrow Cr_j(A) = Pl_j(A)$

# Fonctions de communalité
# Fonctions d'implicabilité



Fonctions à des fins calculatoires de $2^{\Theta}$ à valeurs dans [0,1]

Fonction de communalité

$$Q_j(A) = \sum_{B \subseteq D, B \supseteq A} m_j(B)$$

Fonction d'implicabilité

$$b_j(A) = Cr_j(A) + m_j(\emptyset) = \sum_{B \subseteq A} m_j(B)$$

# Affaiblissement



## Modélisation de la fiabilité

$m'_j(A)=\alpha_j\, m_j(A)$ pour tout $A \neq \Theta$

$m'_j(\Theta)=1-\alpha_j(1-m_j(\Theta))$

$\alpha_j \in [0,1]$ coefficient d'affaiblissement modélise la fiabilité de la source $S_j$ ($\alpha_j=0$ source pas fiable du tout, toute la masse est affectée à $\Theta$ ce qui représente l'ignorance totale)

On augmente les intervalles $[Cr_j, Pl_j]$ (et réduit le conflit lors de la phase de combinaison)

# Pour tenir compte d'une information sûre

Si $A$ de $2^\Theta$ est toujours vérifiée, pour tout $B$ de $2^\Theta$ :

$$m_j(B|A) = \sum_{C \cup A = B} m_j(C) / \sum_{C \cup A \neq \emptyset} m_j(C)$$

ou encore

$$Pl_j(B|A) = \frac{Pl_j(B \cap A)}{Pl_j(A)}$$

Raffinement
Grossissement



**Pour des cadres de discernement différents, mais compatibles**

Ex : $\Theta_1$ = {avion, hélicoptère, missile} et
$\Theta_2$ ={Rafale, Mirage, …, Puma, Gazelle, …, Crotale, Mistral,…}

*Cas des classes incluses de classifieurs non supervisés*

$R$ : fonction de raffinement, pour tout $A$ de $2^{\Theta_1}$

$m_2(R(A))=m_1(A)$

Grossissement $R^{-1}$

# Estimation

Problème : estimer les fonctions de masse

Le choix de la fonction de masse relève de la modélisation et de l'application

- Fonction à support simple
- Fonction sur les singletons
- Méthodes fondées sur les modèles probabilistes
- Méthodes fondées sur les modèles de distances

## Fonction à support simple



Toute la masse de la source $S_j$ porte sur un sous-ensemble non vide $A$ de $2^{\Theta}$ et sur l'ensemble de discernement $\Theta$ (connaissance incertaine et imprécise)

$m_j(A)=s$

$m_j(\Theta)=1-s \quad s\in[0,1]$

$m_j(B)=0$ pour tout $B\neq A$, $B\neq \Theta$

Si $s=0$, alors $m_j(\Theta)=1$ ceci représente l'ignorance totale

# Fonction sur les singletons



La fonction de masse d'une source $S_j$ est définie sur les singletons uniquement

$m_j(\{C_i\})(x)= M_i^j(x)$

$m_j(A)=0$ si $A \notin \{\{C_1\}, \{C_2\},\ldots, \{C_n\}\}$

$M_i^j(x)$ peut être estimée comme une probabilité

- Approche très réductrice, n'utilise pas les propriétés des fonctions de masse
- Beaucoup partent de ces fonctions, puis répartissent les masses sur les hypothèses composées

# Modèles probabilistes



- Ces modèles nécessitent l'estimation de $p(f_j(x)|x \in C_i)$ que l'on notera $p(S_j \mid C_i)$
- 2 modèles proposés par Appriou

  Idée : étendre le modèle sur singletons à l'aide des fonctions simples et en intégrant un coefficient de fiabilité (affaiblissement) de la source pour une classe donnée $\alpha_{ij}$
- 1 modèle proposé par Dromigny-Badin

  Idée : étendre le modèle sur singletons en comparant les probabilités $p(S_j \mid C_i)$

2 modèles reposent sur 2 axiomes

1. Les $n*m$ couples $[M_i^j,\alpha_{ij}]$ sont des sources d'information distinctes dont les éléments focaux sont $\{C_i\}$, $\{C_i\}^c$ et $\Theta$.

2. Si $M_i^j=0$ et $M_i^j$ est valide ($\alpha_{ij}=1$) alors il est certain que $C_i$ n'est pas vérifiée.

- 1. implique le calcul de $m_j^i$ fonctions de masse à partir des $n*m$ couples $[M_i^j,\alpha_{ij}]$
- 2. limite le nombre de fonctions de masse à 2 types

L'axiome 2 entraîne 2 modèles :

Modèle 1 : $m_j^i(\{C_i\})(\mathrm{x}) = M_i^j$

$m_j^i(\{C_i\}^c)(\mathrm{x}) = 1 - M_i^j$

Modèle 2 : $m_j^i(\Theta)(\mathrm{x}) = M_i^j$

$m_j^i(\{C_i\}^c)(\mathrm{x}) = 1 - M_i^j$

Ajout de $\alpha_{ij}$ sous forme d'affaiblissement

Modèle 1 : $m_j^i(\{C_i\})(x) = \alpha_{ij} M_i^j$

$m_j^i(\{C_i\}^c)(x) = \alpha_{ij}(1 - M_i^j)$

$m_j^i(\Theta)(x) = 1 - \alpha_{ij}$

Modèle 2 : $m_j^i(\{C_i\})(x) = 0$

$m_j^i(\{C_i\}^c)(x) = \alpha_{ij}(1 - M_i^j)$

$m_j^i(\Theta)(x) = 1 - \alpha_{ij}(1 - M_i^j)$

Comment déterminer $M_i^j$ ?

(Connaissant la règle de combinaison)

3ème axiome :

3. Conformité avec l'approche bayésienne (cas où $p(S_j|C_i)$ représente parfaitement la réalité ($\alpha_{ij}$ =1 pour tout $i$, $j$) et toutes les probabilités *a priori* $p(\{C_i\})$ sont disponibles)

Les sources $S_j$ étant indépendantes, les résultats doivent être les mêmes si l'on calcule individuellement $p(S_j \mid C_i)$ ou directement la probabilité conjointe $p(S_1,\ldots, S_m \mid C_i)$

Modèle 1 : $M_i^j = \dfrac{R_j\, p(S_j \mid C_i)}{1 + R_j\, p(S_j \mid C_i)}$

avec $R_j$ facteur de normalisation tel que $R_j \geq 0$

Modèle 2 : $M_i^j = R_j\, p(S_j \mid C_i)$

avec $R_j \in [0, (\max_{s_j,i}(p(s_j / C_i)))^{-1}]$

On obtient les fonctions de masse

Modèle 1 :

$$m_j^i\left(\{C_i\}\right)(x) = \frac{\alpha_{ij} R_j\, p(S_j \mid C_i)}{1 + R_j\, p(S_j \mid C_i)}$$

$$m_j^i\left(\{C_i\}^c\right)(x) = \frac{\alpha_{ij}}{1 + R_j\, p(S_j \mid C_i)}$$

$$m_j^i(\Theta)(x) = 1 - \alpha_{ij}$$

avec $R_j \geq 0$

Modèle 2 :

$$m_j^i(\{C_i\})(x) = 0$$

$$m_j^i(\{C_i\}^c)(x) = \alpha_{ij}\,(\,1 - R_j\,p(S_j \mid C_i)\,)$$

$$m_j^i(\Theta)(x) = 1 - \alpha_{ij}\,(\,1 - R_j\,p(S_j \mid C_i)\,)$$

avec $R_j \in [0, (\max_{s_j, i}(p(s_j / C_i)))^{-1}]$

- $\alpha_{ij}$ : coefficient d'affaiblissement fixé proche de 1 et $p(S_j/C_i)$ peut être estimé à partir des matrices de confusion

- Adaptés aux cas où l'on apprend une classe contre toutes les autres

Cas où $\{C_i\}$ et $\Theta$ sont les seuls éléments focaux, on suppose connu $\boldsymbol{x}_i$ un prototype (centre) de $\{C_i\}$

$m_j^i(\{C_i\})(x) = \alpha_{ij} \exp[-\gamma_{ij}\, d^2(x, \boldsymbol{x}_i)]$

$m_j^i(\Theta)(x) = 1 - \alpha_{ij} \exp[-\gamma_{ij}\, d^2(x, \boldsymbol{x}_i)]$

- $0<\alpha_{ij}<1$ est un coefficient d'affaiblissement et $\gamma_{ij}>0$, ce sont des paramètres pour jouer sur la quantité d'ignorance et de la forme des fonctions de masse
- La distance permet d'affecter une masse à $x$ plus importante en fonction de sa ressemblance à $\{C_i\}$

# Appriou *vs* Denœux



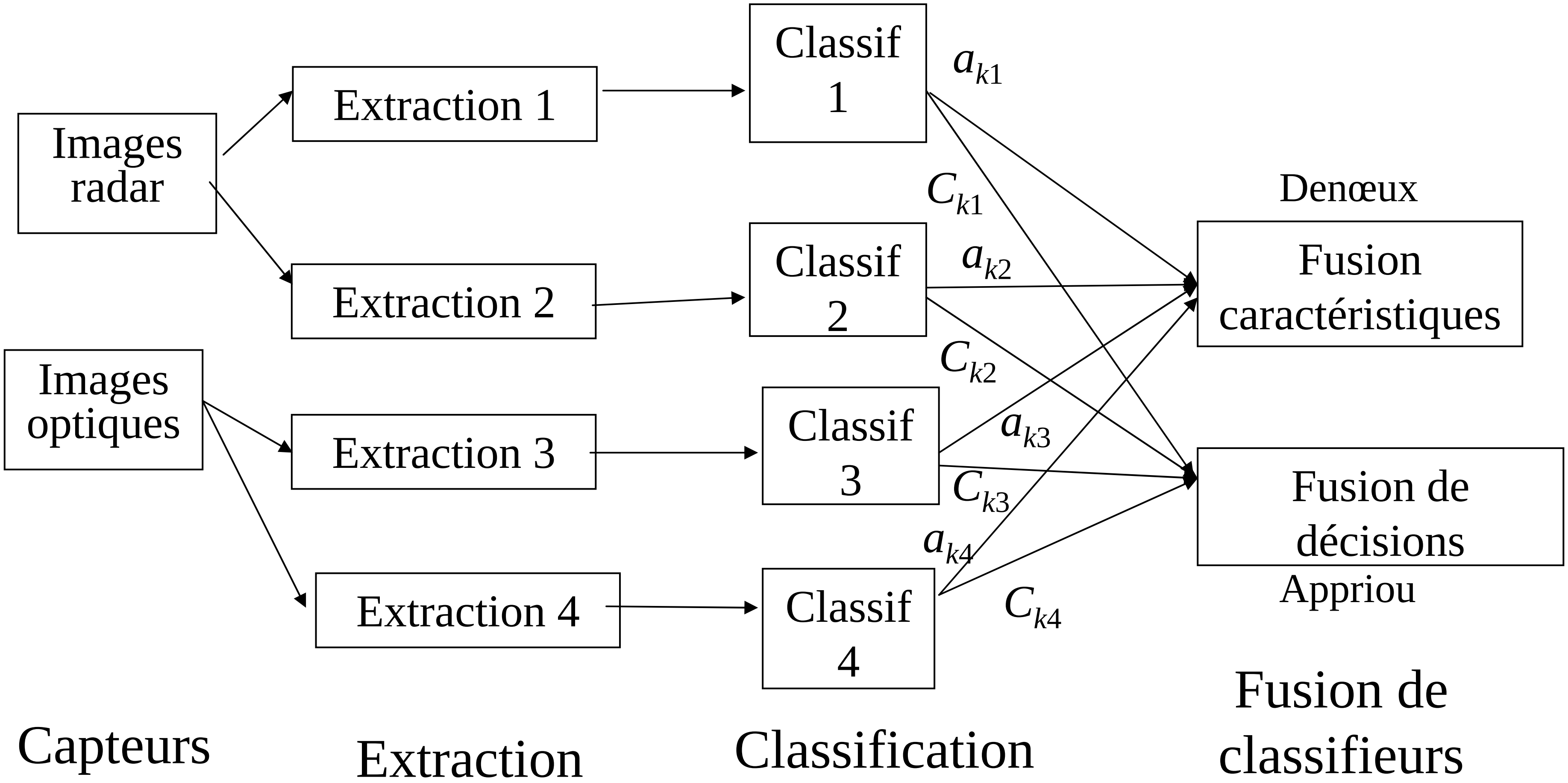
- Appriou : apprentissage de $p(S_j/C_i)$
- Denœux : choix de la distance $\mathrm{d}(x, \boldsymbol{x}_i)$

- $p(S_j/C_i)$ plus facile à estimer sur les décisions des classifieurs par la matrice de confusion
- $\mathrm{d}(x, \boldsymbol{x}_i)$ plus facile à choisir sur les sorties numériques des classifieurs (ex : distance euclidienne)

Fusion de décisions
Fusion de caractéristiques

Exemple de fusion

# Fusion de Classifieurs



# Fusion de classifieurs

$$\begin{array}{ccc}
 & \begin{matrix} C_1 & \cdots & C_i & \cdots & C_n \end{matrix} & \\
\begin{matrix} S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{matrix} &
\begin{bmatrix}
m_1^1(x) & \cdots & m_i^1(x) & \cdots & m_n^1(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
m_1^j(x) & \cdots & m_i^j(x) & \cdots & m_n^j(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
m_1^m(x) & \cdots & m_i^m(x) & \cdots & m_n^m(x)
\end{bmatrix} &
\begin{matrix} m_1 \\ \vdots \\ m_j \\ \vdots \\ m_m \end{matrix}
\end{array}
\xrightarrow{\text{Fusion}} \text{La classe}$$

# Fusion de classifieurs

$$
\begin{array}{cc}
 & \begin{array}{ccccc} C_1 & \cdots & C_i & \cdots & C_n \end{array} \\
\begin{array}{c} S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{array} &
\left[\begin{array}{ccccc}
m_1^1(x) & \cdots & \boxed{m_i^1(x)} & \cdots & m_n^1(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
m_1^j(x) & \cdots & m_i^j(x) & \cdots & m_n^j(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
m_1^m(x) & \cdots & m_i^m(x) & \cdots & m_n^m(x)
\end{array}\right] \\
 & \begin{array}{ccccc} m_1 & \cdots & m_i & \cdots & m_n \end{array}
\end{array}
$$

$$\Big\downarrow \text{ Fusion}$$

La classe

# Fusion de classifieurs

$$
\begin{array}{cc}
 & \begin{array}{ccccc} C_1 & \cdots & C_i & \cdots & C_n \end{array} \\
\begin{array}{c} S_1 \\ \vdots \\ S_j \\ \vdots \\ S_m \end{array} &
\begin{bmatrix}
m_1^1(x) & \cdots & m_i^1(x) & \cdots & m_n^1(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
m_1^j(x) & \cdots & m_i^j(x) & \cdots & m_n^j(x) \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
m_1^m(x) & \cdots & m_i^m(x) & \cdots & m_n^m(x)
\end{bmatrix}
\end{array}
$$

$$\xrightarrow{\text{Combinaison}} m \xrightarrow{\text{Décision}} \text{La classe}$$

# Combinaison

- Combinaison conjonctive (intersection)
- Combinaison disjonctive (union)
- Combinaison mixte

Règle orthogonale de Dempster - Shafer pour $A$ de $2^{\Theta}$

$$m(A) = (m_1 \oplus m_2 \oplus \ldots \oplus m_m)(A) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_m = A} \prod_{j=1}^{m} m_j(B_j)$$

Permet de synthétiser toutes les fonctions de masses $m_j^i$

S'écrit simplement avec les fonctions de communalité $Q(A) = \sqcap_{j=1}^{m} Q_j(A)$

- La combinaison de 2 experts est donnée par :

$$m(A) = \sum_{B_1 \cap B_2 = A} m_1(B_1) m_2(B_2)$$

- Exemple

| | $\emptyset$ | $A$ | $B$ | $C$ | $\Theta$ |
|---|---|---|---|---|---|
| $m_1$ | 0 | 0.5 | 0.1 | 0 | 0.4 |
| $m_2$ | 0 | 0.2 | 0 | 0.5 | 0.3 |
| $m$ | 0.32 | 0.33 | 0.03 | 0.2 | 0.12 |

- Problème : masse non nulle sur l'ensemble vide
  - en monde ouvert : représente une solution non prévue
  - en monde fermé : pas acceptable

Conjonctive | Normalisation | Combinaison 4/15

$$m(A) = \frac{\displaystyle\sum_{B_1 \cap B_2 \cap \ldots \cap B_m = A} \prod_{j=1}^{m} m_j(B_j)}{1-k} \quad \text{si } A \neq \emptyset$$

et $m(\emptyset)=0$

avec $k = \displaystyle\sum_{B_1 \cap B_2 \cap \ldots \cap B_m = \emptyset} \prod_{j=1}^{m} m_j(B_j) < 1$

$k$ est une mesure de conflit entre les sources (dépend de la modélisation) (ou encore *inconsistance de la fusion*)

## Origines du conflit

– Les sources ne sont pas fiables. L'information est erronée et peut conduire à une ambiguïté

– Le cadre de discernement est non exhaustif. Hypothèse d'un monde fermé fausse

– Les sources observent des phénomènes différents. Dans ce cas il ne faut pas les combiner

## Problème : la normalisation masque le conflit (Exemple de Zadeh)

# Exemple de Zadeh

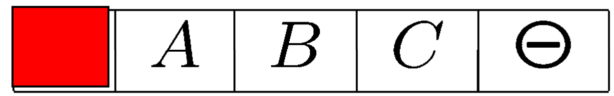
| | $\emptyset$ | $A$ | $B$ | $C$ | $\Theta$ |
|---|---|---|---|---|---|
| $m_1$ | 0 | 0.9 | 0 | 0.1 | 0 |
| $m_2$ | 0 | 0 | 0.9 | 0.1 | 0 |
| $m_{12}$ normalisée | 0 | 0 | 0 | 1 | 0 |
| $m_{12}$ non normalisée | 0.99 | 0 | 0 | 0.01 | 0 |

Toute la masse est sur $C$ seul élément où les 2 sources sont d'accord.

Problème la normalisation masque le conflit

Intéressant si monde fermé sans conflit

TBF (*Transferable Belief* Functions) en monde ouvert




$$m(A) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_m = A \neq \emptyset} \prod_{j=1}^{m} m_j(B_j)$$

$$m(\emptyset) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_m = \emptyset} \prod_{j=1}^{m} m_j(B_j)$$

*k* est affecté à l'ensemble vide

Yager propose un modèle en monde fermé où la mesure de conflit est affectée au cadre de discernement total $\Theta$

| $\emptyset$ | $A$ | $B$ | $C$ | |
|---|---|---|---|---|

$$m(A)=\sum_{B_1\cap B_2\cap\ldots\cap B_m=A\neq\emptyset}\prod_{j=1}^{m}m_j(B_j)$$

Calcul de $m(\Theta)$ :

$$m(\Theta)=1-\sum_{A\subseteq D}\sum_{B_1\cap B_2\cap\ldots\cap B_m=A\neq\emptyset}\prod_{j=1}^{m}m_j(B_j)$$

puis la modifie

$$m(\Theta)=m(\Theta)+\sum_{B_1\cap B_2\cap\ldots\cap B_m=\emptyset}\prod_{j=1}^{m}m_j(B_j)$$

La technique du *hedging* consiste a ajouter un élément inconnu $e$ au cadre de discernement *i.e.* que l'on suppose que le conflit vient du manque d'exhaustivité

| | | | | | |
|---|---|---|---|---|---|
| $\emptyset$ | $A$ | $B$ | $C$ | | $\ominus$ |

$$m(A \cup e) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_m = A \neq \emptyset} \prod_{j=1}^{m} m_j(B_j)$$

$$m(e) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_m = \emptyset} \prod_{j=1}^{m} m_j(B_j)$$

- La règle de combinaison de Dempster (normalisée ou non) est commutative et associative
- La fonction de masse définie par

$m_0(\Theta)=1$ et $m_0(A)=0$ pour tout $A \subseteq \Theta$ et $A \neq \Theta$

est l'élément neutre de la combinaison

- Règle applicable sous l'hypothèse d'indépendance cognitive des sources

- La loi $\oplus$ n'est pas idempotente

| | $\emptyset$ | $A$ | $B$ | $C$ | $\Theta$ |
|---|---|---|---|---|---|
| $m_1$ | 0 | 0.7 | 0.2 | 0.1 | 0 |
| $m_2$ | 0 | 0.7 | 0.2 | 0.1 | 0 |
| $m_{12}$ normalisée | 0 | 0.91 | 0.07 | 0.02 | 0 |
| $m_{12}$ non normalisée | 0.46 | 0.49 | 0.04 | 0.01 | 0 |

Le conflit de deux sources identiques est non nul !

Auto-conflit

- Réduit l'imprécision et l'ambiguïté

Exemple 1 :

| $\cap$ | $A \cup B$ | $C$ |
|---|---|---|
| $A$ | $A$ | $\emptyset$ |
| $B \cup C$ | $B$ | $C$ |

Résout l'ambiguïté et réduit l'imprécision

Exemple 2 :

| $\cap$ | $A \cup B$ | $C$ | $\ominus$ |
|---|---|---|---|
| $A$ | $A$ | $\emptyset$ | $A$ |
| $B \cup C$ | $B$ | $C$ | $B \cup C$ |
| $\ominus$ | $A \cup B$ | $C$ | $\ominus$ |

Ambiguïté partiellement réduit $B \cup C,\ A \cup B$

- Perte du conflit
- Élargit les éléments focaux donc perte de spécificité

$$m(A) = \sum_{B_1 \cup B_2 \cup \ldots \cup B_m = A} \prod_{j=1}^{m} m_j(B_j)$$

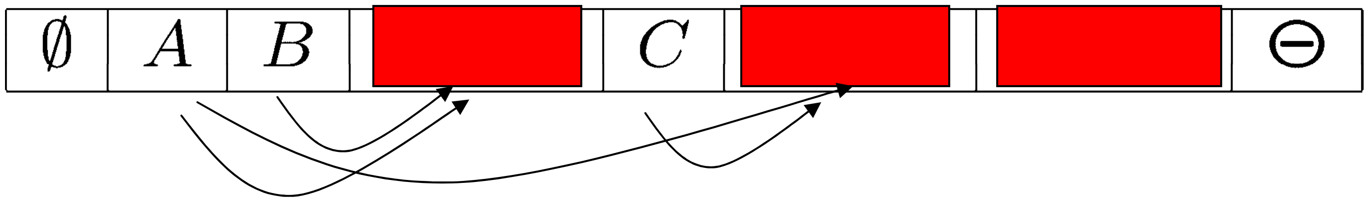
Intéressant si on ne sait pas modéliser les fiabilités des sources, leurs ambiguïtés et imprécisions.

$$b(A) = \sqcap_{j=1}^{m} b_j(A)$$

Compromis des approches conjonctives et disjonctives

$$m(A) = \sum_{B_1 \cap B_2 \cap \ldots \cap B_m = A} \prod_{j=1}^{m} m_j(B_j) + \sum_{\substack{B_1 \cup B_2 \cup \ldots \cup B_m = A \\ B_1 \cap B_2 \cap \ldots \cap B_m = \emptyset}} \prod_{j=1}^{m} m_j(B_j)$$

Répartition fine du conflit



## Combinaison conjonctive

$$m(A) = \sum_{B_1 \cap B_2 = A} m_1(B_1) m_2(B_2)$$

**en redistribuant le conflit** sur les singletons

on ajoute :

$m_1^2(A)m_2(B)/(m_1(A)+m_2(B))$ à $m(A)$ et

$m_1(A)m_2^2(B)/(m_1(A)+m_2(B))$ à $m(B)$

| | $\emptyset$ | $A$ | $B$ | $\Theta$ |
|---|---|---|---|---|
| $m_1$ | 0 | 0.6 | 0 | 0.4 |
| $m_2$ | 0 | 0.3 | 0.2 | 0.5 |
| $m_c$ | 0.12 | 0.6 | 0.08 | 0.2 |
| $m_{PCR6}$ | 0 | 0.69 | 0.11 | 0.2 |

# Décision

- Maximum de plausibilité
- Maximum de crédibilité
- Maximum de crédibilité avec rejet
- Maximum de probabilité pignistique

# Maximum de plausibilité



$$x \in C_i \text{ si } Pl(\{C_i\})(x) = \max\{Pl(\{C_k\})(x),\ 1 \leq k \leq n\}$$

Optimal pour des fonctions de masses dérivées de probabilité (cohérent avec le maximum de probabilité *a posteriori*)

Problème : manque parfois de pouvoir discriminant (trop optimiste)

# Maximum de crédibilité

Décision 3/5

$$x \in C_i \ \text{ si } Cr(\{C_i\})(x) = \max\{Cr(\{C_k\})(x),\ 1 \le k \le n\}$$

Identique au critère du maximum de plausibilité dans le cas où le résultat de la combinaison ne porte que sur des singletons (les éléments focaux ne sont pas nécessairement réduits aux singletons dans ce cas)

Plus sélective, pessimiste

Problème : utilisée seule peut entraîner beaucoup d'erreurs

# Maximum de crédibilité avec rejet

Décision 4/5

$$x \in C_i \text{ si } Cr(\{C_i\})(x) = \max\{Cr(\{C_k\})(x),\ 1 \leq k \leq n\}$$
$$\text{et } Cr(\{C_i\})(x) \geq Cr(\{C_i\}^c)(x)$$

La décision n'est prise que si elle est suffisamment non ambiguë (on doit croire plus en $\{C_i\}$ qu'en son contraire)

Plus restrictif que le critère du maximum de crédibilité donc plus discriminant

Problème : peut conduire à aucune décision

# Maximum de probabilité pignistique

Décision 5/5

Probabilité pignistique (Smets) projette les masses définies sur $2^{\Theta}$ les singletons

$$betP(X) = \sum_{Y \in 2^{\Theta}, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m(Y)}{1 - m(\emptyset)}.$$

où $|A|$ est le cardinal de $A$.

Décision probabiliste

$$x \in C_i \text{ si } bet(\{C_i\})(x) = \max\{bet(\{C_k\})(x), 1 \leq k \leq n\}$$

- Bien dans le cas d'ambiguïtés, prudent
- Peu conforme à la notion de masse

# Fonctions de croyance pour l'association

- Appliqué en suivi de cibles
- Principe :

Un cadre de discernement pour l'association cible/piste et un cadre pour l'association piste/cible

- $\Theta = \{Y_1, Y_2, ..., Y_n, *\}$

où $Y_j$ correspond à l'association avec $Y_j$

et * correspond à l'hypothèse d'exhaustivité (association à rien de connu)

# Interprétation des fonctions de masse — Association

- $m_{i,j}(Y_j)$ masse associée à « $X_i$ est reliée à $Y_j$ »
- $m_{j,i}(X_i)$ masse associée à « $Y_j$ est reliée à $X_i$ »
- $m_{i,j}(\overline{Y_j})$ masse associée à « $X_i$ n'est pas reliée à $Y_j$ »
- $m_{i,j}(\Theta)$ masse associée à « On ignore si $X_i$ est reliée à $Y_j$ »
- $m_{i,.}(*)$ masse associée à « $X_i$ est reliée à rien »

# Pour la classification non supervisée

**Association**

- Un cadre de discernement pour chaque association $S_i/S_j$ de classifieurs
- $\Theta_{ij} = \{C_1, C_2, ..., C_{n_j}, *\}$

où $C_l$ correspond à l'association d'une classe du classifieur $S_i$ avec la classe $C_l$ du classifieur $S_j$

et * correspond à l'hypothèse d'exhaustivité (association d'une classe du classifieur $S_i$ à aucune classe connue du classifieur $S_j$)

- On obtient des matrices de masse pour chaque couple $(S_i,S_j)$ de classifieurs

# Pour la classification non supervisée

**Association**

- $m_{i,j}(C_l)$ distribution de masses associée à la relation des classes du classifieur $S_i$ avec la classe $C_l$ du classifieur $S_j$
- $m_{j,i}(C_k)$ distribution de masses associée à la relation des classes du classifieur $S_j$ avec la classe $C_k$ du classifieur $S_i$
- $m_{i,j}(\overline{C_l})$ distribution de masses associée à la non-relation des classes du classifieur $S_i$ avec la classe $C_l$ du classifieur $S_j$
- $m_{i,j}(\Theta)$ distribution de masses associée à l'ignorance sur la relation des classes du classifieur $S_i$ avec une classe (ou disjonction) du classifieur $S_j$
- $m_{i,.}(*)$ distribution de masses associée à la relation des classes du classifieur $S_i$ avec aucune classe du classifieur $S_j$

# Pour la classification non supervisée

## Association

- La théorie des fonctions de croyance offre la possibilité d'associer $C_k$ du classifieur $S_i$ à $\{C_{l_1} \cup C_{l_2}\}$ ou encore à $\{C_{l_1} \cap C_{l_2}\}$ dans le cadre de la DSmT

# Réflexions

- Limites
  - complexité à croissance exponentielle
  - indépendance cognitives des sources
  - l'estimation des fonctions de masse passe souvent par une phase d'apprentissage et dépend de l'application

- Liens avec la théorie des probabilités et la théorie des possibilités

# Liens avec la théorie des probabilités

Réflexion

Lorsque les seuls éléments focaux sont des singletons

- les fonctions de masses sont des probabilités
- $Cr_j$ et $P_j$ sont égales aux fonctions de masses
- la loi de combinaison de Dempster est cohérente avec les lois des probabilités
- Les probabilités sont un cas particulier de la théorie des fonctions de croyance

# Liens avec la théorie des possibilités

Réflexion

Lorsque les éléments focaux sont tous inclus les uns dans les autres

– les fonctions de crédibilité sont les fonctions de nécessité dans la théorie des possibilités

– les fonctions de plausibilité sont les fonctions de possibilités dans la théorie des possibilités

Donc les notions de probabilité et de possibilité sont incompatibles (sauf si il n’y a qu’un élément focal)