Processus d'agrégation pour la décision

Alain Appriou

ONERA
BP 72, 29 avenue de la Division Leclerc
92322 Châtillon Cedex
Alain.Appriou@onera.fr

Résumé. La robustesse, l'acuité, et la réactivité d'un système repose le plus souvent sur l'utilisation conjointe de moyens d'observation complémentaires, et donc sur l'agrégation d'informations disparates en termes d'incertitude, d'imprécision, d'incomplétude, de fiabilité, de subjectivité, ou de pertinence. La théorie de l'évidence constitue alors un cadre propice au développement d'un ensemble d'opérateurs cohérents, propres à traiter les problèmes de modélisation unifiée d'informations variées, d'association de données ambiguës, de combinaison de sources disparates, conflictuelles, et dépendantes, et de prise de décision. La chaîne de traitement complète et modulable qui peut être constituée à l'aide de ces opérateurs est également applicable à la décision multicritère, qu'elle soit implantée au niveau décisionnel d'un système, qu'elle concerne la gestion de ses ressources, ou qu'elle contribue à sa conception.

1 Systèmes complexes et fusion de données

La conception de systèmes complexes revient le plus souvent à définir la meilleure association de moyens, susceptible d'assurer la mise en œuvre d'actions particulières à partir d'observations appropriées. Ces observations sont en général issues de sources multiples, suffisamment complémentaires pour répondre aux exigences les plus sévères en matière de robustesse au contexte et à l'environnement, d'acuité et de richesse de l'information exploitée, et de réactivité opérationnelle. L'évaluation de tels systèmes requiert une parfaite connaissance des modules de traitement retenus pour la fusion des informations et l'aide à la décision qu'ils doivent assurer, ces éléments ayant un impact majeur sur les performances de l'ensemble du système.

L'élaboration d'un processus de fusion des données exige cependant une bonne maîtrise des difficultés liées, d'une part à la disparité des informations traitées, que ce soit en termes d'incertitude, d'imprécision, d'incomplétude, de fiabilité, de subjectivité, ou de pertinence, et d'autre part à la complexité des tâches d'analyse et à la sûreté des décisions qui doivent être mises en œuvre. Un certain nombre de fonctionnalités propres aux traitements de fusion exigent en particulier des développements spécifiques :

- la modélisation conjointe et l'intégration dans un même formalisme d'informations de nature différente, pouvant chacune relever d'approches théoriques variées : observations, connaissances a priori, apprentissages, fiabilités ;
- l'association de données ambiguës, dans l'espace ou dans le temps ;

- la combinaison de sources disparates : prise en compte de référentiels distincts, intégration d'informations contextuelles, gestion de conflits, exploitation des dépendances, traitement dynamique, asynchronisme ;
- les principes, les critères, et les niveaux de décision, le processus devant en outre avoir la capacité de gérer un ensemble de décisions les plus pertinentes, compte tenu de l'information disponible.

Au-delà de ces besoins, d'autres préoccupations apparaissent au niveau de l'ingénierie du système lui-même, en particulier :

- l'élaboration des architectures de traitement les mieux adaptées, que ce soit en termes de niveau de fusion, de centralisation ou de distribution, de hiérarchisation, ou d'organisation fonctionnelle ;
- la définition de la gestion des ressources : critères d'affectation, « orthogonalisation » des sources, optimisation dynamique, coopération de moyens autonomes, adaptativité à l'environnement ;
- les principes de conception du système, avec la modélisation des contraintes et des préférences telles qu'elles peuvent être appréhendées, l'évaluation suffisamment pertinente des performances *a priori*, le processus de décision pour le choix de la solution adoptée, et la méthodologie d'optimisation face à la diversité des degrés de liberté.

La mise en œuvre, le plus souvent conjointe, d'un certain nombre de cadres théoriques appropriés, permet de surmonter ces difficultés de façon circonstanciée. Parmi les théories les plus utiles, la théorie de l'évidence occupe une place privilégiée, tant par son caractère fédérateur que par la pertinence des opérateurs qu'elle permet de concevoir. Il est en particulier possible de développer dans le cadre de cette théorie, à partir d'une même notion de base, un ensemble cohérent d'opérateurs propres à traiter les différents problèmes évoqués. Ces opérateurs permettent de constituer une chaîne de traitement complète, modulable, capable d'appréhender globalement les difficultés mises en évidence.

Cette approche peut également être utilisée de façon duale pour apporter une réponse circonstanciée aux besoins en décision multicritère qui peuvent apparaître, tant au niveau des processus décision implantés au sein du système, qu'au niveau de la gestion de ses ressources, ou surtout pour ce qui concerne sa conception.

2 Notions de base

La théorie de l'évidence (Shafer 1976) repose sur la définition d'un cadre de discernement E, ensemble de I hypothèses exclusives et exhaustives H_i ($i \in [1,I]$). 2^E désigne alors l'ensemble des 2^I sous-ensembles de E, incluant l'ensemble vide.

2.1 Fonctions élémentaires

Trois fonctions, définies de 2^E sur l'intervalle [0,1], permettent de caractériser la vraisemblance de chacun des sous-ensembles de E :

- la fonction de masse $m(.)$, qui représente la vraisemblance de chacun des singletons H_i appartenant au sous-ensemble de E en argument, sans qu'il soit néanmoins possible de distinguer entre ces singletons ; elle est telle que :

$$\sum_{A \subseteq E} m(A) = 1 \qquad [1]$$

$$m(\emptyset) = 0 \qquad [2]$$

- la fonction de croyance $Cr(.)$, qui peut être interprétée comme la vraisemblance minimale du sous-ensemble de E en argument, et qui est définie de façon unique à partir d'une fonction de masse donnée :

$$Cr(B) = \sum_{A \subseteq B} m(A) \qquad [3]$$

- la fonction de plausibilité $Pl(.)$, qui peut être interprétée comme la vraisemblance maximale du sous-ensemble E en argument, et qui est elle aussi reliée de façon biunivoque aux fonctions de masse et de croyance, respectivement par les relations :

$$Pl(B) = \sum_{A \cap B \neq \emptyset} m(A) \qquad [4]$$

$$Pl(B) = 1 - Cr(\neg B) \qquad [5]$$

où $\neg B$ désigne le complémentaire de B dans E.

Les éléments focaux d'une fonction de masse $m(.)$ sont par ailleurs les éléments A de 2^E tels que $m(A)$ n'est pas nul.

2.2 Conditionnement

Le conditionnement permet de prendre en compte une information selon laquelle il est certain que le sous-ensemble A de E est vérifié. Il transforme en conséquence une fonction de masse quelconque $m(.)$ sur E en une fonction de masse $m(./\mathrm{A})$ dont tous les éléments focaux sont inclus dans A, ceci en transférant la masse de chaque élément focal de $m(.)$ sur sa partie incluse dans A, et en renormalisant par rapport à la masse qui a ainsi pu être affectée :

$$m(B \,/\, A) = \frac{\sum_{C \cap A = B} m(C)}{\sum_{C \cap A \neq \emptyset} m(C)} \qquad [6]$$

Cette expression peut également s'écrire à l'aide des fonctions de plausibilité, pour tout sous-ensemble B du cadre de discernement E :

$$Pl(B \,/\, A) = \frac{Pl(B \cap A)}{Pl(A)} \qquad [7]$$

Le déconditionnement est l'opération inverse du conditionnement. Connaissant une fonction de masse $m(./A)$ sur un sous-ensemble A de E, le but est de reconstituer une fonction de masse complète $m(.)$ sur E. L'expression [6] met clairement en évidence l'indétermination du problème si l'on ne connaît que $m(./A)$, le nombre d'inconnues étant alors nécessairement supérieur au nombre d'équations. Il convient donc d'utiliser toute connaissance complémentaire disponible, et au-delà de rechercher la fonction $m(.)$ de spécificité minimale, c'est-à-dire qui affecte au maximum la masse non définie aux éléments focaux les plus grands possibles.

De façon plus pragmatique, l'inversion de [7] fournit directement la plausibilité de tous les sous-ensembles de A :

$$\forall B \subseteq A \quad Pl(B) = Pl(B \,/\, A)\, Pl(A) \qquad [8]$$

Outre $Pl(./A)$, la reconstruction complète de $m(.)$ exige donc d'estimer $Pl(\mathrm{A})$, ainsi que la plausibilité de tous les sous-ensembles de E non inclus dans A. Dans le cas général où ces grandeurs sont totalement inconnues, le minimum de spécificité déjà évoqué nous conduit à les prendre égales à 1. Ceci définit entièrement la fonction de masse $m(.)$ ainsi obtenue par l'opérateur de déconditionnement classique :

$$\forall B \subseteq A \quad m(B \cup \neg A) = m(B / A) \qquad [9]$$

2.3 Raffinement, grossissement

Un raffinement R associe à chaque hypothèse H_i^1 d'un cadre de discernement $E^1=\{H_1^1,\ldots,H_{I1}^1\}$ un sous-ensemble $R(H_i^1)$ d'un autre cadre de discernement $E^2=\{H_1^2,\ldots,H_{I2}^2\}$ tel que $\{R(H_1^1),\ldots,R(H_{I1}^1)\}$ constitue une partition de E^2. L'opération de raffinement revient donc à considérer que chaque singleton H_i^1 de E^1 est lui-même représentatif d'un ensemble d'hypothèses plus détaillées $R(H_i^1)$. Une fonction de masse $m^1(.)$ définie sur E^1 fournit donc naturellement, par une opération dite d'extension minimale, une fonction de masse $m^2(.)$ sur E^2 :

$$\forall A \subseteq E^1 \quad m^2(R(A)) = m^1(A) \qquad [10]$$

Le grossissement est l'opération inverse R^{-1} du raffinement R. Il consiste donc à regrouper les singletons du cadre de discernement E^2 en sous-ensembles exclusifs $R(H_i^1)$, qui sont alors associés aux singletons H_i^1 du cadre de discernement E^1. Une fonction de masse $m^2(.)$ définie sur E^2 conduit dans ces conditions à une fonction de masse $m^1(.)$ sur E^1 par la transformation :

$$m^1(A) = \sum_{\substack{B \subseteq E^2 \\ A=\{H_i^1 / R(H_i^1) \cap B \neq \emptyset\}}} m^2(B) \qquad [11]$$

3 Gestion de la dépendance des observations

Une grandeur discriminante X appartenant à un cadre de discernement $E_1=\{X_1,\ldots,X_{N1}\}$ et une grandeur discriminante Y appartenant à un cadre de discernement $E_2=\{Y_1,\ldots,Y_{N2}\}$ peuvent décrire une même situation observée selon deux caractéristiques complémentaires, par exemple la vitesse d'une cible pour l'un et la taille de cette cible pour l'autre. Ces deux grandeurs n'ont aucune raison d'être indépendantes, dans le sens où toutes les associations de vitesses et de tailles ne sont pas nécessairement réalisables, ou du moins pas avec la même vraisemblance. La connaissance que l'on peut avoir sur les dépendances entre les grandeurs X et Y peut être prise en compte en évaluant directement sur le produit cartésien $E_1 \times E_2$ la fonction de masse $m_{12}(.)$ représentative de l'ensemble de la connaissance disponible.

Ceci peut se faire en exprimant sur $E_1 \times E_2$ la fonction de plausibilité $Pl_{12}(.)$ associée à $m_{12}(.)$, pour tout sous-ensemble A de E_1 et tout sous-ensemble B de E_2 (Appriou 2002) :

$$Pl_{12}(A \times B) = Pl_1(A / B \subseteq E_2)\, Pl_2(B) \qquad [12]$$

Typiquement l'évaluation $Pl_2(B)$ de Y sur E_2 peut résulter d'une observation, et l'évaluation conditionnelle $Pl_1(A/B \subseteq E_2)$ de X sur E_1 peut provenir d'une connaissance *a priori*.

Les fonctions de plausibilité $Pl_2(B)$ et $Pl_1(A/B \subseteq E_2)$ peuvent n'être pas entièrement définies. De plus [12] ne permet pas d'exprimer tous les termes de $Pl_{12}(.)$ sur $E_1 \times E_2$. La fonction $Pl_{12}(.)$ peut néanmoins toujours être complétée par minimisation de la spécificité, c'est-à-dire en donnant la plus grande valeur possible aux plausibilités inconnues, de façon à ce que $Pl_{12}(.)$ ne représente bien que la connaissance disponible.

Il est par ailleurs évident que E_1 et E_2 peuvent jouer des rôles symétriques dans [12]. En particulier, si les deux senseurs donnent respectivement une évaluation de X sur E_1 et une évaluation de Y sur E_2, chacune d'elles doit donner lieu à une fonction de plausibilité sur $E_1 \times E_2$ via [12]. Les fonctions de plausibilité ainsi obtenues peuvent ensuite être combinées par une des règles qui vont être introduites au paragraphe 4.

Il convient enfin de remarquer que si les grandeurs X et Y sont indépendantes [12] devient :

$$Pl_{12}(A \times B) = Pl_1(A)\, Pl_2(B) \qquad [13]$$

4 Combinaison d'observations

Le résultat précédent permet de proposer une formulation générale de la combinaison de sources distinctes. Supposons que deux grandeurs X et Y soit évaluées sur les cadres de discernement E_1 et E_2 introduits au paragraphe 3, et que l'on cherche à en déduire l'évaluation d'une grandeur Z appartenant à un cadre de discernement $E_3=\{Z_1,\ldots,Z_{N3}\}$, moyennant la connaissance préalable de relations entre la grandeur Z et les grandeurs X et Y. Typiquement X et Y sont toujours les grandeurs discriminantes invoquées au paragraphe 3, et Z est une classe de situations pour laquelle les grandeurs X et Y possibles sont connues de façon plus ou moins sûre.

Le résultat [12] peut alors être appliqué au produit cartésien $E_1 \times E_2 \times E_3$, fournissant pour tout sous-ensemble A de E_1, tout sous-ensemble B de E_2, et tout sous-ensemble C de E_3 :

$$Pl_{123}(A \times B \times C) = Pl_3(C / A \times B \subseteq E_1 \times E_2)\, Pl_{12}(A \times B) \qquad [14]$$

Si les observations constituent des sources distinctes, $Pl_{12}(A \times B)$ est bien sûr donnée par [13]. $Pl_3(C/A \times B)$ formalise la connaissance que l'on a sur les relations susceptibles d'unir l'ensemble de classes C aux ensembles A et B de grandeurs discriminantes. Il ne reste plus éventuellement qu'à conditionner $Pl_{123}(.)$ sur les singletons (X_i, Y_h, Z_k) admissibles pour le problème traité, et à effectuer un grossissement de $E_1 \times E_2 \times E_3$ sur E_3 pour obtenir l'évaluation cherchée.

Ce formalisme général de combinaison peut être particularisé pour retrouver toutes les règles de combinaison de sources connues à ce jour (Appriou 2002). En particulier, parmi les plus usitées :

- si $E_1=E_2=E_3$, $Pl_3(C/A \times B)=1$ pour tout A, B, et C, et les singletons (X_i, Y_h, Z_k) admissibles sont tels que $X_i=Y_h=Z_k$, alors la procédure correspond à la somme orthogonale :

$$m(C) = \frac{\sum\limits_{A \cap B = C \neq \emptyset} m_1(A)\, m_2(B)}{1 - \sum\limits_{A \cap B = \emptyset} m_1(A)\, m_2(B)} \qquad [15]$$

- si $E_1=E_2=E_3$, $Pl_3(C/A \times B)=1$ pour $C \cap (A \cup B) \neq \emptyset$, $Pl_3(C/A \times B)=0$ pour $C \cap (A \cup B) = \emptyset$, et tous les singletons (X_i, Y_h, Z_k) sont admissibles, alors la procédure correspond à la règle disjonctive :

$$m(C) = \sum_{A \cup B = C} m_1(A)\, m_2(B) \qquad [16]$$

- si $E_1=E_2=E_3$, $Pl_3(C/A \times B)=1$ pour $C \cap A \cap B \neq \emptyset$, $Pl_3(C/A \times B)=1$ pour $A \cap B = \emptyset$ et $C \cap (A \cup B) \neq \emptyset$, $Pl_3(C/A \times B)=0$ pour $A \cap B \neq \emptyset$ et $C \cap A \cap B = \emptyset$, $Pl_3(C/A \times B)=0$ pour

$A \cap B = \emptyset$ et $C \cap (A \cup B) = \emptyset$, et tous les singletons (X_i, Y_h, Z_k) sont admissibles, alors la procédure correspond à la combinaison mixte de Dubois et Prade :

$$m(C) = \sum_{A \cap B = C} m_1(A)\, m_2(B) + \sum_{\substack{A \cup B = C \\ A \cap B = \emptyset}} m_1(A)\, m_2(B) \quad [17]$$

5 Modélisation des données

Considérons l'ensemble E composé de I hypothèses H_i ($i \in [1,I]$), exclusives et exhaustives, parmi lesquelles il convient de choisir celle qui correspond le mieux à la situation en cours d'analyse. Par ailleurs, J grandeurs u_j ($j \in [1,J]$), pertinentes pour ce problème de discrimination sur E, font chacune l'objet d'une observation s_j.

Chaque observation s_j peut être utilisée pour fournir, sur la base d'une caractérisation *a priori* de chaque hypothèse H_i, I fonctions de coût C_{ij} prenant valeur dans [0,1], chacune représentative de la vraisemblance de l'hypothèse H_i correspondante. Un facteur de fiabilité q_{ij} à valeur dans [0,1] est également associé à chaque vraisemblance C_{ij}. Il a pour but d'exprimer l'aptitude de cette vraisemblance C_{ij} à discriminer l'hypothèse H_i, compte tenu de la qualité de la connaissance disponible. Il peut en particulier, par exemple, figurer la plus ou moins bonne représentativité de l'apprentissage *a priori* utilisé pour élaborer C_{ij}, en regard du contexte opérationnel rencontré. Il est typiquement élaboré à partir de l'observation o_{ij} d'une variable contextuelle z_{ij}, capable de discrimination sur le cadre de discernement $E_{ij}=\{F_{ij}, \neg F_{ij}\}$, où F_{ij} est la proposition : « C_{ij} est fiable ».

L'observation s_j d'une grandeur discriminante u_j ou l'observation o_{ij} d'une variable contextuelle z_{ij} peuvent typiquement être une valeur exacte, une observation incertaine caractérisée par une densité de probabilité conditionnelle $p(s_j/u_j)$ (resp. $p(o_{ij}/z_{ij})$), ou une observation imprécise fournissant un sous-ensemble flou $\mu_j(u_j)$ des valeurs de u_j possibles (resp. $\mu_{ij}(z_{ij})$).

De même la connaissance *a priori* qui nous permet de caractériser, d'une part les valeurs de la grandeur u_j pour l'hypothèse H_i, et d'autre part les valeurs de z_{ij} pour lesquelles nous sommes sûrs que C_{ij} est fiable, peut revêtir la forme de valeurs déterministes u_{ij} (resp. z_{Fij}) bien identifiées, de distributions $p(u_j/H_i)$ (resp. $p(z_{ij}/F_{ij})$) obtenues par apprentissage statistique, ou de fonctions d'appartenance $\mu_i(u_j)$ (resp. $\mu_{Fij}(z_{ij})$) caractérisant une connaissance imprécise.

Pour interpréter C_{ij} et de q_{ij} dans le cadre de la théorie de l'évidence, les problèmes couramment traités sont formalisés sur la base de deux axiomes (Appriou 2002) :

Axiome 1 : Chacun des $I \times J$ couples $[C_{ij}, q_{ij}]$ constitue une source d'information distincte, dont les éléments focaux sont H_i, $\neg H_i$, et E. De plus C_{ij} représente l'évidence placée sur H_i et sur toutes les hypothèses susceptibles de lui ressembler, compte tenu des caractéristiques fournies par l'observation s_j. De même, $1-C_{ij}$ représente l'évidence placée sur toutes les hypothèses appartenant à $\neg H_i$ et sur toute hypothèse susceptible de leur ressembler, dans les mêmes conditions d'analyse.

Axiome 2 : Si $C_{ij}=0$ et C_{ij} est valide ($q_{ij}=1$), alors il est certain que H_i n'est pas vérifiée.

Ceci conduit à interpréter chacune des $I \times J$ vraisemblances C_{ij} à l'aide d'une fonction de masse $m_{Vij}(.)$ qui peut dès lors revêtir deux formes distinctes. La première est définie par :

$$m_{Vij}(\neg H_i) = 1 - C_{ij} \quad [18]$$

$$m_{Vij}(E) = C_{ij} \quad [19]$$

La deuxième fonction de masse possible est :

$$m_{Vij}(H_i) = C_{ij} \quad [20]$$

$$m_{Vij}(\neg H_i) = 1 - C_{ij} \quad [21]$$

Corrélativement, l'interprétation de chacun des $I \times J$ facteurs de fiabilité q_{ij} à l'aide d'une fonction de masse $m_{Fij}(.)$ sur E_{ij} conduit elle aussi à deux formulations, la première étant :

$$m_{Fij}(\neg F_{ij}) = 1 - q_{ij} \quad [22]$$

$$m_{Fij}(E_{ij}) = q_{ij} \quad [23]$$

et la deuxième :

$$m_{Fij}(F_{ij}) = q_{ij} \quad [24]$$

$$m_{Fij}(\neg F_{ij}) = 1 - q_{ij} \quad [25]$$

La parfaite symétrie des problèmes et des formulations entre l'élaboration des $m_{Vij}(.)$ et celle des $m_{Fij}(.)$ conduit à des expressions strictement identiques de C_{ij} et de q_{ij} en fonction des observations et des apprentissages, respectivement des grandeurs discriminantes et des variables contextuelles. Ces expressions peuvent être obtenues pour toutes les combinaisons possibles de natures entre les observations et les apprentissages, telles qu'elles ont pu être introduites précédemment (Appriou 2002).

La synthèse entre les fonctions de masse $m_{Vij}(.)$ et $m_{Fij}(.)$ sous la forme d'une fonction de masse unique $m_{ij}(.)$ est effectuée en considérant que la plausibilité associée à $m_{Vij}(.)$ doit être vue comme la plausibilité de tout sous-ensemble A de E, conditionnellement au fait que la source $m_{Vij}(.)$ est fiable, soit $Pl_{Vij}(A/F_{ij} \in E_{ij})$. Il convient donc de déterminer sur $E\text{x}E_{ij}$ la fonction de masse $m'_{ij}(.)$ de minimum de spécificité qui satisfait, par application de [12] :

$$Pl'_{ij}(A \times F_{ij}) = Pl_{Vij}(A / F_{ij} \in E_{ij})\, Pl_{Fij}(F_{ij}) \quad [26]$$

et dont le grossissement de $E\text{x}E_{ij}$ sur E_{ij} est la fonction $m_{Fij}(.)$. Le grossissement de $m'_{ij}(.)$ de $E\text{x}E_{ij}$ sur E fournit finalement la fonction de masse cherchée $m_{ij}(.)$. Compte tenu des différentes formulations possibles pour $m_{Vij}(.)$ et pour $m_{Fij}(.)$, $m_{ij}(.)$ peut être donnée par deux modèles. Le *modèle* 1 s'exprime :

$$m_{ij}(H_i) = 0 \quad [27]$$

$$m_{ij}(\neg H_i) = q_{ij}\,(1 - C_{ij}) \quad [28]$$

$$m_{ij}(E) = 1 - q_{ij}\,(1 - C_{ij}) \quad [29]$$

et le *modèle* 2 :

$$m_{ij}(H_i) = q_{ij}\, C_{ij} \quad [30]$$

$$m_{ij}(\neg H_i) = q_{ij}\,(1 - C_{ij}) \quad [31]$$

$$m_{ij}(E) = 1 - q_{ij} \quad [32]$$

Une fonction de masse $m(.)$ synthétisant l'ensemble des évaluations peut être obtenue en combinant les fonctions de masse $m_{ij}(.)$ de la façon la plus informative, c'est-à-dire par somme orthogonale. Il convient de noter que le *modèle* 1 est celui qui minimise la spécificité.

Par ailleurs, cette approche de la modélisation permet de résoudre plus simplement et plus efficacement certains problèmes (Appriou 2002). En particulier, si l'information est incomplète, c'est-à-dire si un ou plusieurs C_{ij} ne sont pas accessibles par manque d'apprentissage ou d'observabilité des grandeurs discriminantes concernées, il suffit d'ignorer la ou les fonctions de masse $m_{ij}(.)$ correspondantes. Le déconditionnement peut également être traité de façon plus informative au niveau des $m_{ij}(.)$ plutôt qu'au niveau de $m(.)$, compte tenu de la définition particulière des éléments focaux de $m_{ij}(.)$, sachant que le conditionnement et la somme orthogonale peuvent être appliqués dans n'importe quel ordre. De même la connaissance de relations de compatibilité entre des hypothèses, c'est-à-dire de leur ressemblance perçue par certaines grandeurs discriminantes, peut conduire à un déconditionnement plus fin si elle est prise en compte au niveau des fonctions de masse $m_{ij}(.)$.

6 Décision : choix des hypothèses les plus vraisemblables

Plutôt que de chercher de façon classique, mais souvent abusive, l'hypothèse unique la plus vraisemblable, l'objectif est ici de sélectionner le sous-ensemble A de E qui a le plus de chances de contenir la bonne hypothèse. Afin de ne pas déclarer systématiquement le cadre de discernement total, il convient que le sous-ensemble A choisi réalise également un compromis avec un critère de cardinal $|A|$ minimal.

En conséquence, nous devons considérer comme entrées du processus de décision, d'une part la fonction de masse $m(.)$ établie sur E à partir des observations, et d'autre part une fonction de masse $m_d(.)$ définie sur $E_d=\{d_A\}$, où d_A représente le choix respectivement de chaque sous-ensemble A de E. Cette fonction de masse $m_d(.)$ doit rassembler tous les souhaits *a priori* concernant la décision escomptée. C'est donc typiquement une fonction de masse Bayésienne exprimée sur E_d par :

$$m_d(d_A) = K_d\ \lambda_A\ g(|A|) \qquad [33]$$

La fonction $g(.)$ est une fonction monotone décroissante quelconque du cardinal du sous-ensemble A évalué, qui permet d'éviter de déclarer l'ensemble E dans sa totalité en ajustant un compromis entre la certitude et la spécificité de l'ensemble des hypothèses retenues. Elle peut préférentiellement revêtir la forme :

$$g(|A|) = \left(\frac{1}{|A|}\right)^r \qquad [34]$$

où r est un paramètre appartenant à [0,1] qui permet de choisir une attitude parmi un *continuum* de principes de décision allant du choix d'un singleton (r=1) à l'indécision totale (r=0).

Le coefficient λ_A intègre le manque de connaissance sur l'une quelconque des hypothèses appartenant à A, afin d'éviter de choisir systématiquement les hypothèses sur lesquelles peu d'information est disponible. La constante K_d est un facteur de normalisation qui assure la compatibilité avec la notion de fonction de masse.

Le traitement de $m(.)$ et $m_d(.)$ consiste à raffiner ces deux fonctions sur ExE_d, à les combiner sur cet espace par somme orthogonale, à conditionner le résultat sur les éléments (H_i,d_A) de $E\times E_d$ tels que $H_i\in A$, puis à effectuer un grossissement de $E\times E_d$ dans E_d . Ceci fournit une fonction de masse Bayésienne sur E_d, dont le maximum procure la décision optimale d_A* :

$$d_A{}^* = \arg\left(\max_{A\subset E}\left(m_d(d_A)\,Pl(A)\right)\right) \qquad [35]$$

où $m_d(.)$ est bien sûr donnée par [33].

Il convient de noter que l'application de cette loi de décision aux modèles proposés au paragraphe 5 conduit à une formulation analytique globale relativement simple, et donc à une mise en œuvre rapide et peu coûteuse, en particulier pour le *modèle* 1 (Appriou 2002).

Par ailleurs, cette procédure peut conduire à des ambiguïtés de décision qui doivent prioritairement faire l'objet d'une recherche paramétrique de la solution de cardinal supérieur le plus proche qui ne soit plus ambiguë. Si cette démarche ne s'avère pas satisfaisante, les solutions équivalentes obtenues peuvent être départagées en cherchant parmi elles le sous-ensemble A dont le complémentaire est de surcroît le moins vraisemblable :

$$d_A{}^* = \arg\left(\min_{A\subset E} (m_d(d_{\neg A})\,Pl(\neg A))\right) \qquad [36]$$

7 Mise en œuvre des opérateurs dégagés

La cohérence mise en évidence entre les différents éléments qui viennent d'être présentés conduit naturellement à les intégrer ensemble pour constituer une chaîne modulaire complète de traitement, de l'interprétation des données à la prise de décision, en vue de gérer de façon circonstanciée la plupart des difficultés rencontrées en pratique, au-delà de ce que permettent les approches plus classiques. Cet avantage est néanmoins le plus souvent obtenu au prix d'un accroissement plus ou moins important de la charge de traitement, qui doit donc être limitée en cherchant à mettre directement en œuvre la formulation analytique équivalente à l'ensemble de la chaîne de traitement.

8 Application à la décision multicritère

Le processus ainsi élaboré peut également être appliqué de façon relativement immédiate à tout problème de décision multicritère, pour apporter une réponse adaptée aux besoins qui peuvent apparaître, tant au niveau des processus de décision implantés au sein du système, qu'au niveau de la gestion de ses ressources, ou surtout pour ce qui concerne sa conception. L'idée est alors de rechercher l'action la plus vraisemblablement susceptible de satisfaire les attentes l'utilisateur, et ceci pour les contextes opérationnels qu'il envisage. Il convient dans ces conditions de considérer que :

- les hypothèses H_i du cadre de discernement sont maintenant les différentes actions possibles,
- les grandeurs discriminantes u_j deviennent les indices de performance qui vont permettre de mesurer le niveau de satisfaction de l'utilisateur pour chaque action possible. Les souhaits s_j de l'utilisateur peuvent en effet être exprimés de façon imprécise sur ces échelles de préférence (comme un sous-ensemble flou de valeurs défini par $\mu_j(u_j)$), ou de façon incertaine (à l'aide de distributions de valeurs de la propriété considérée $p(s_j/u_j)$). Le crédit accordé à chaque action possible H_i est alors déterminé à partir des valeurs d'indice attendues, et de la connaissance plus ou moins imprécise (sous-ensembles flous de valeurs défini par $\mu_i(u_j)$), ou incertaine (distribution $p(u_j/H_i)$) que l'on peut avoir *a priori* sur les performances obtenues grâce à la mise en œuvre de l'action H_i considérée.
- Corrélativement, les conditions de mise en œuvre o_{ij} envisagées par l'opérateur pour chaque action H_i sont pour leur part exprimées en fonction de variables contextuelles z_{ij}, de façon imprécise (sous-ensembles flous $\mu_{ij}(z_{ij})$) ou incertaine (distributions $p(o_{ij}/z_{ij})$). La connaissance *a priori* que l'on a sur la validité de l'évaluation concernant les performances de l'action H_i peut elle aussi être exprimée sur l'espace

de ces variables contextuelles de façon imprécise (sous-ensembles flous $\mu_{Fij}(z_{ij})$) ou incertaine (distributions $p(z_{ij}/F_{ij})$).

La parfaite similitude entre le problème de choix multicritère ainsi posé et celui traité au paragraphe 5 permet d'utiliser directement la modélisation qui y est élaborée, et par suite tous les outils décrits dans cet article afin de mettre en œuvre la chaîne complète de traitement proposée, jusqu'à la procédure de décision qui permet de dégager l'ensemble des actions les plus pertinentes.

9 Conclusion

Le bénéfice de cette approche a déjà pu être exploité dans un certain nombre d'applications relatives à la classification de cibles. Un atout majeur du processus de traitement réside dans sa convivialité avec les principaux cadres théoriques couramment utilisés pour traiter les données numériques. Tout en reposant sur une formulation analytique simple, il permet en particulier de gérer de façon rigoureuse :

- la variété de nature des informations, des connaissances, ou des souhaits ;
- l'absence partielle d'information ;
- les référentiels d'analyse et les relations mal définies entre eux ;
- un ensemble de décisions les plus pertinentes, et en particulier ses compromis vraisemblance/précision ;
- les ambiguïtés de décision.

Les évolutions attendues aujourd'hui en matière de traitement concernent principalement l'intégration cohérente de l'ensemble des techniques envisageables, notamment en développant les synergies possibles entre d'une part l'approche présentée ici, et d'autre part les logiques propres à traiter le niveau symbolique, les méthodes de raisonnement, et l'intelligence distribuée dans les architectures réparties et reconfigurables.

Pour les aspects système, les voies d'investigation doivent maintenant porter de façon privilégiée sur le caractère dynamique de la gestion des moyens, sur l'évaluation globale des performances *a priori*, et sur le processus de recherche des solutions en conception, avec une attention toute particulière pour ses facultés de traçabilité.

Références

Appriou A., Discrimination multisignal par la théorie de l'évidence, dans Reconnaissance des formes et décision en signal, Traité IC2, Information, Commande, Communication, Editions Hermès, 2002.

Shafer G., A Mathematical Theory of Evidence, Princeton University Press, Princeton, New Jersey, 1976.

Summary

The robustness, the efficiency, and the reactivity of a system, most often requires the joint use of complementary observation sources, and therefore the aggregation of disparate information as regards uncertainty, imprecision, incompleteness, reliability, subjectivity, or relevance. The theory of evidence provides a suitable framework for the development of coherent operators that use a common formalism to deal with heterogeneous information

modelling, ambiguous data matching, disparate, conflicting, and dependent source combination, and decision making. These operators allow the elaboration of a complete and modular process, that is also suitable for multicriterion decision making. The latter arises especially at the decision level of an autonomous system, at the resource management level, or for the system design.