

Optimisation entropique des systèmes de détection distribuée

Entropy optimisation for distributed detection systems

par **Christophe DESROUSSEAU** et **Denis POMORSKI**

Laboratoire d'Automatique et d'Informatique Industrielle de Lille, UPRES A 8021,
Université des Sciences et Technologies de Lille, Cité Scientifique, Bâtiment P2
59655 Villeneuve d'Ascq cedex, France
e-mail : denis.pomorski@univ-lille1.fr

résumé et mots clés

Dans le contexte général de la fusion de données, la théorie de la détection permet de discriminer de façon optimale, à partir d'observations, un nombre fini de situations. Classiquement, les critères de Bayes et de Neyman-Pearson sont utilisés dans la littérature.

Pourtant, dans certaines applications (en particulier lorsque les probabilités *a priori* P_j , $j = 0, 1$ des hypothèses sont mal connues, ou si l'une d'entre elles est proche de zéro), il pourrait s'avérer intéressant de baser le problème d'optimisation sur la quantité d'information pertinente pour le problème de détection. Cette approche pourrait notamment s'adapter aux problèmes de communication numérique où l'on s'intéresse davantage à la quantité d'information transmise, plutôt qu'à l'information elle-même. Pour de tels systèmes, un critère entropique peut être intéressant.

En utilisant un tel critère, nous rappelons les résultats de l'optimisation entropique des architectures de détection centralisée. Dans le cadre de la détection décentralisée parallèle, nous proposons d'optimiser les détecteurs locaux sans prendre en compte le problème de la fusion, que nous optimisons séparément. L'intérêt principal de cette approche est qu'elle offre une solution simple à mettre en œuvre. Nous rappelons alors les résultats de l'optimisation simultanée des détecteurs locaux et de l'opérateur de fusion. Nous proposons également d'utiliser le critère informationnel dans le cadre de l'optimisation d'une architecture décentralisée série.

Enfin, une comparaison des résultats de l'optimisation de toutes ces architectures sur un exemple illustratif nous permet de mettre en évidence des propriétés très intéressantes de robustesse vis-à-vis d'événements rares de l'approche informationnelle, contrairement à l'approche classique.

Fusion de données, détection décentralisée, optimisation, entropie de Shannon

abstract and key words

In the general framework of data fusion, detection theory allows to distinguish a finite set of situations from observations. Bayes and Neyman-Pearson criteria are generally used in order to develop optimal detection rules.

Nevertheless, in some applications (in particular when the *a priori* probabilities P_j , $j = 0, 1$ of the hypotheses are uncertain or when one of them is near zero), it could be interesting to consider the optimisation of the quantity of information which is relevant for the detection problem. This approach could be used in digital communication problems where the information quantity sent is more interesting than the information itself. For such systems, an entropy based criterion seems to be of interest.

Using such a criterion, we recall the entropy optimisation results for centralised detection networks. Within the framework of parallel distributed detection, we suggest to optimise the local detectors without taking into account the fusion problem, that we optimize separately. The main interest of this approach is that its implementation is simple. We recall then the results of the simultaneous optimisation of the local detectors and of the fusion operator. We also use an informational criterion for the optimisation of serial distributed networks.

Finally, the different optimisation results are compared on an illustrative example. This allows to underline the very interesting robustness properties with respect to rare events of the informational approach, as opposed to the classical one.

Data fusion, distributed detection, optimisation, Shannon's entropy.

1. introduction

La théorie de la détection permet de discriminer de façon optimale, à partir d'observations, un nombre fini de situations.

La détection d'un signal à partir de plusieurs capteurs peut être envisagée de deux manières différentes. Traditionnellement, l'ensemble des capteurs communique directement leurs observations au détecteur central qui prend la décision finale. Cette approche, appelée « détection centralisée », nécessite souvent des lignes de communication à large bande si l'on veut obtenir une décision en temps réel. La deuxième approche, appelée « détection décentralisée », consiste à associer à chaque capteur un détecteur qui décide localement si un signal a été ou non détecté. Les décisions locales sont ensuite envoyées à un opérateur de fusion qui les combine afin de prendre la décision finale. L'avantage de cette approche est la réduction des coûts de communication. En contrepartie, les performances obtenues sont moins bonnes puisque l'opérateur de fusion ne reçoit pas toute l'information nécessaire à la prise de décision finale.

Dans la littérature, les critères de Bayes et de Neyman-Pearson [1] sont très utilisés afin d'optimiser les systèmes de détection centralisée et décentralisée. Dans le cas Bayésien, un coût est associé à chaque situation et une fonction « risque moyen » (ou probabilité d'erreur) est minimisée. Dans les applications pour lesquelles ces coûts sont connus et ont une signification précise, l'approche Bayésienne est une excellente solution au problème de décision. Ce n'est cependant pas forcément le cas pour toutes les applications et une solution différente serait l'utilisation du critère de Neyman-Pearson qui consiste à chercher la meilleure probabilité de détection pour une probabilité de fausse alarme donnée.

Pourtant, dans certaines applications (notamment lorsque les probabilités *a priori* P_j $j = 0, 1$ des hypothèses sont mal connues, ou si l'une d'entre elles est proche de zéro), il pourrait s'avérer intéressant de baser le problème d'optimisation sur la quantité d'information pertinente pour le problème de détection. Cette approche pourrait notamment s'adapter aux problèmes de communication numérique où l'on s'intéresse davantage à la quantité d'information transmise, plutôt qu'à l'information elle-même. Pour de tels systèmes, un critère entropique pourrait être plus approprié. Différents travaux ont été effectués dans ce sens; Middleton [2] et Gabrielle [3] ont proposé une architecture de détection centralisée basée sur la minimisation de la perte d'information entre l'entrée et la sortie du système. Hoballah [4] et Warren [5] ont étendu ce résultat à l'optimisation d'une architecture décentralisée parallèle.

L'idée intuitive de l'utilisation de l'entropie en détection provient de l'analogie entre les systèmes de détection et les systèmes de transmission binaire. Dans le cadre des systèmes de transmission binaire, une information H prenant la valeur H_0 ou H_1 est envoyée *via* un canal de transmission. Le récepteur reçoit un bit u_0 qui représente l'information H perturbée (figure 1). Suivant

les valeurs prises par H et par u_0 , on répertorie, comme dans le cadre de la détection, quatre situations. A chacune de ces situations, on associe une probabilité qui peut s'exprimer en fonction des probabilités de fausse alarme P_F et de détection P_D du système (figure 2). L'objectif est alors de minimiser la perte d'information sur H (H_0 ou H_1) connaissant la décision u_0 ($u_0=0$ ou $u_0=1$). Cette perte d'information peut être mesurée par l'entropie conditionnelle de Shannon $h(H/u_0)$, qui représente l'incertitude sur H connaissant la décision u_0 [6].

Nous proposons tout naturellement, comme critère d'optimisation des systèmes de détection, la minimisation de cette entropie conditionnelle.

Ce travail est inspiré des travaux de Hoballah et Varshney [4] [1] développant une approche informationnelle du problème de la détection distribuée basée sur la maximisation de la transinformation interne $i(H, u_0)$ [7]¹. Nous rappelons les résultats de l'optimisation entropique des architectures de détection centralisée. Dans le cadre de la détection décentralisée parallèle, nous proposons d'optimiser les détecteurs locaux sans prendre en compte le problème de la fusion, que nous optimisons séparément. L'intérêt principal de cette approche est qu'elle offre une solution simple à mettre en œuvre. Nous rappelons alors les résultats de l'optimisation simultanée des détecteurs locaux et de l'opérateur de fusion. Nous utilisons également le critère informationnel dans le cadre de l'optimisation d'une architecture décentralisée série.

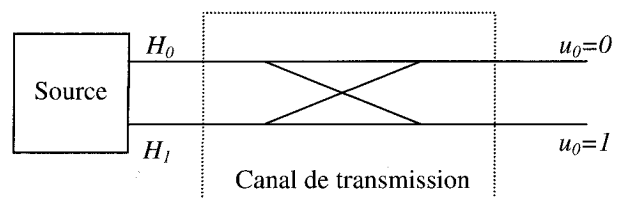


Figure 1. – Principe de fonctionnement d'un système de transmission binaire.

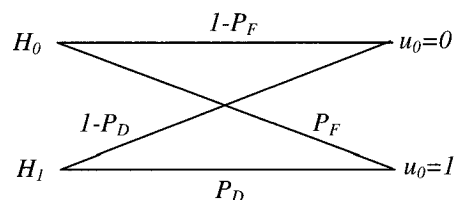


Figure 2. – Modèle de transmission appliqué au problème de la détection.

1. Critère équivalent à celui que nous proposons car

$$i(H, u_0) = h(H) - h(H/u_0) \text{ et } h(H) = - \sum_{j=0}^1 P_j \cdot \log P_j = C^{te}.$$

Enfin, une comparaison des résultats de l'optimisation de toutes ces architectures sur un exemple illustratif nous permet de mettre en évidence des propriétés très intéressantes de robustesse vis-à-vis d'événements rares de l'approche informationnelle, contrairement à l'approche classique.

2. utilisation d'un critère entropique dans le cadre de la détection centralisée

Considérons le système de la figure 3 composé de N capteurs Y_i ($i = 1, \dots, N$) observant le même phénomène. A partir des mesures $y = (y_1, y_2, \dots, y_N)$ fournies par le vecteur des capteurs $Y = (Y_1, Y_2, \dots, Y_N)$, une décision u_0 ($u_0 = 0$ ou $u_0 = 1$) doit être prise. Les lois de probabilités *a priori* P_j ($j = 0, 1$) associées à chaque hypothèse sont supposées connues, et on note :

- la probabilité de détection $P_D = P(u_0 = 1/H_1)$
- la probabilité de fausse alarme $P_F = P(u_0 = 1/H_0)$

Le critère entropique consiste à minimiser l'entropie conditionnelle de Shannon $h(H/u_0)$ entre H et u_0 [2] [3]. Dans un premier temps, nous nous attachons à déterminer l'expression de l'entropie conditionnelle $h(H/u_0)$ en fonction des probabilités de détection P_D et de fausse alarme P_F . Nous étudions dès lors les variations de cette fonction et démontrons les relations existant entre le critère entropique et les critères de Bayes et de Neyman-Pearson.

2.1. expression de l'entropie conditionnelle $h(H/u_0)$ en fonction de P_D et de P_F

L'entropie conditionnelle $h(H/u_0)$ s'écrit :

$$h(H/u_0) = - \sum_{i,j \in \{0,1\}} P(u_0 = i, H_j) \cdot \log P(H_j/u_0 = i) \quad (1)$$

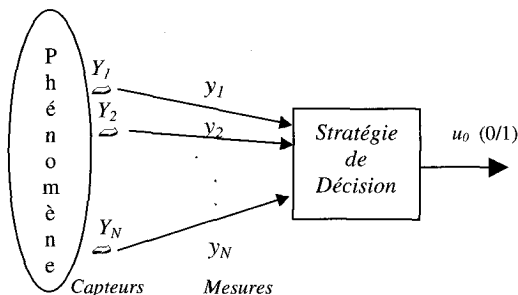


Figure 3. – L'architecture de détection centralisée.

$$h(H/u_0) = - \sum_{i,j \in \{0,1\}} P(u_0 = i/H_j) \cdot P_j \cdot \log \frac{P(u_0 = i/H_j) \cdot P_j}{P(u_0 = i/H_0) \cdot P_0 + P(u_0 = i/H_1) \cdot P_1} \quad (2)$$

On en déduit que :

$$\begin{aligned} h(H/u_0) &= -P_0(1 - P_F) \log \frac{P_0(1 - P_F)}{P_0(1 - P_F) + (1 - P_0)(1 - P_D)} \\ &\quad - P_0 P_F \log \frac{P_0 P_F}{P_0 P_F + (1 - P_0) P_D} \\ &\quad - (1 - P_0)(1 - P_D) \log \frac{(1 - P_0)(1 - P_D)}{P_0(1 - P_F) + (1 - P_0)(1 - P_D)} \\ &\quad - (1 - P_0) P_D \log \frac{(1 - P_0) P_D}{P_0 P_F + (1 - P_0) P_D} \end{aligned} \quad (3)$$

L'entropie $h(H/u_0)$ s'exprime donc en fonction des probabilités de fausse alarme P_F , de détection P_D , et de P_0 .

En notant :

$$\begin{aligned} \alpha_1 &= P_0 P_F & \beta_1 &= (1 - P_0) P_D \\ \alpha_{-1} &= P_0(1 - P_F) & \beta_{-1} &= (1 - P_0)(1 - P_D) \end{aligned} \quad (4)$$

l'équation (3) s'écrit :

$$h(H/u_0) = - \sum_{i \in \{-1,1\}} \left[\alpha_i \log \frac{\alpha_i}{\alpha_i + \beta_i} + \beta_i \log \frac{\beta_i}{\alpha_i + \beta_i} \right] \quad (5)$$

2.2. étude des variations de $h(H/u_0)$

Attachons-nous maintenant à l'étude de la fonction $h(H/u_0) = f(P_D, P_F)$ à P_0 constante. Pour cela, nous étudions les variations de cette fonction à P_F constante, puis à P_D constante.

a) Etude des variations de $h(H/u_0)$ à P_F constante

Notons qu'à P_F constante, les relations suivantes sont obtenues :

$$\frac{\partial \alpha_i}{\partial P_D} = 0 \quad \text{et} \quad \frac{\partial \beta_i}{\partial P_D} = i \cdot (1 - P_0) \quad i = -1, 1 \quad (6)$$

Nous démontrons en annexe 1 que la dérivée de la fonction $f(v) = \alpha \log \frac{\alpha}{\alpha + \beta} + \beta \log \frac{\beta}{\alpha + \beta}$ par rapport à v peut se mettre sous la forme :

$$\frac{\partial f(v)}{\partial v} = \frac{\partial \alpha}{\partial v} \log \frac{\alpha}{\alpha + \beta} + \frac{\partial \beta}{\partial v} \log \frac{\beta}{\alpha + \beta} \quad (7)$$

Par conséquent, en dérivant $h(H/u_0)$ par rapport à P_D pour une probabilité de fausse alarme P_F fixée, et en utilisant les propriétés (6) et (7), nous obtenons :

$$\begin{aligned} \frac{\partial h(H/u_0)}{\partial P_D} &= -(1 - P_0) \cdot \log \frac{\beta_1}{\alpha_1 + \beta_1} + (1 - P_0) \cdot \log \frac{\beta_{-1}}{\alpha_{-1} + \beta_{-1}} \\ &= (1 - P_0) \cdot \log \frac{(\alpha_1 + \beta_1) \cdot \beta_{-1}}{(\alpha_{-1} + \beta_{-1}) \cdot \beta_1} \end{aligned} \quad (8)$$

De plus,

$$\begin{aligned} \frac{\partial^2 h(H/u_0)}{\partial P_D^2} &= (1 - P_0)^2 \cdot \left[\frac{1}{\alpha_1 + \beta_1} - \frac{1}{\beta_1} + \frac{1}{\alpha_{-1} + \beta_{-1}} - \frac{1}{\beta_{-1}} \right] < 0 \end{aligned} \quad (9)$$

car $\alpha_i + \beta_i > \beta_i \quad i = -1, 1$.

Enfin, $\left(\frac{\partial h(H/u_0)}{\partial P_D} \right)_{P_D=P_F} = 0$.

$\frac{\partial h(H/u_0)}{\partial P_D}(P_D)$ est donc une fonction décroissante de P_D , positive sur $[0, P_F[$ et négative sur $]P_F, 1]$. Par conséquent, $h(H/u_0)(P_D)$ est une fonction croissante sur $[0, P_F[$ et décroissante sur $]P_F, 1]$. En ne prenant en compte que l'intervalle $]P_F, 1]$, c'est-à-dire l'intervalle pour lequel $P_D > P_F$ (seul cas intéressant pour la détection), on en déduit que $h(H/u_0)$ est une fonction décroissante de P_D .

A P_F fixée, chercher le détecteur qui minimise $h(H/u_0)$ est donc équivalent à chercher le détecteur qui maximise la probabilité de détection P_D (figure 4).

b) Etude des variations de $h(H/u_0)$ à P_D constante

Dans l'expression (5), les termes α_i et β_i sont symétriques, et un raisonnement identique au précédent peut être appliqué. Nous

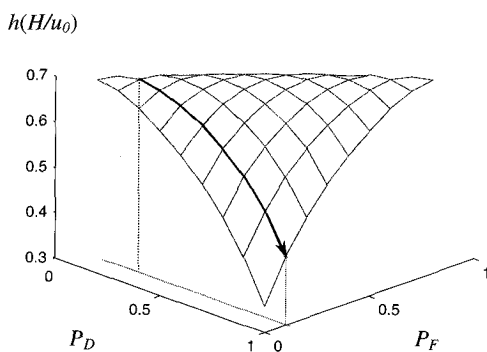


Figure 4. - Minimisation de l'entropie $h(H/u_0)$ à probabilité de fausse alarme P_F constante.

trouvons ainsi que $\frac{\partial^2 h(H/u_0)}{\partial P_F^2} < 0$ à P_D constante. De plus,

$$\left(\frac{\partial h(H/u_0)}{\partial P_F} \right)_{P_F=P_D} = 0.$$

$\frac{\partial h(H/u_0)}{\partial P_F}$ est donc une fonction décroissante de P_F , positive sur $[0, P_D[$ et négative sur $]P_D, 1]$. $h(H/u_0)(P_F)$ est donc croissante sur $[0, P_D[$ et décroissante sur $]P_D, 1]$. Dans le cadre de la détection, on ne retiendra que l'intervalle $[0, P_D[$, c'est-à-dire l'intervalle pour lequel $P_F < P_D$. A P_D fixée, chercher le détecteur qui minimise $h(H/u_0)$ est donc équivalent à chercher le détecteur qui minimise la probabilité de fausse alarme P_F .

En associant les résultats de a) et b), on en conclut que minimiser $h(H/u_0)$ sans fixer P_F ni P_D , revient à résoudre le problème bicritère $\{\min P_F, \max P_D\}$ en adoptant une fonction d'utilité non linéaire $h(H/u_0)$, à l'inverse du critère de Bayes qui minimise la fonction linéaire $R = \alpha \cdot P_F + \beta \cdot (1 - P_D)$.

Montrons alors que minimiser $h(H/u_0)$ revient à effectuer un seuillage du rapport de vraisemblance qui peut être déterminé à partir de la connaissance des probabilités *a posteriori* associées aux différentes situations possibles.

2.3. détermination du seuil λ qui minimise $h(H/u_0)$

Montrons que chercher la valeur du seuil λ pour laquelle l'entropie conditionnelle $h(H/u_0)$ est minimale revient à chercher la valeur du rapport de vraisemblance Λ pour laquelle

$$\left(\frac{\partial h(H/u_0)}{\partial \Lambda} \right)_{\Lambda=\lambda} = 0 :$$

Pour cela, rappelons les équations liant les probabilités de fausse alarme et de détection au rapport de vraisemblance :

$$\begin{aligned} P_F &= \int_{\lambda}^{+\infty} p(\Lambda/H_0) \cdot d\Lambda \\ &\Rightarrow P(\Lambda = \lambda/H_0) = - \left(\frac{\partial P_F}{\partial \Lambda} \right)_{\Lambda=\lambda} \end{aligned} \quad (10)$$

$$\begin{aligned} P_D &= \int_{\lambda}^{+\infty} p(\Lambda/H_1) \cdot d\Lambda \\ &\Rightarrow P(\Lambda = \lambda/H_1) = - \left(\frac{\partial P_D}{\partial \Lambda} \right)_{\Lambda=\lambda} \end{aligned} \quad (11)$$

A partir des équations (4), (10) et (11), les relations suivantes peuvent être établies :

$$\left(\frac{\partial \alpha_i}{\partial \Lambda} \right)_{\Lambda=\lambda} = -i \cdot P_0 \cdot P(\Lambda = \lambda/H_0)$$

et

$$\left(\frac{\partial \beta_i}{\partial \Lambda} \right)_{\Lambda=\lambda} = -i \cdot (1 - P_0) \cdot P(\Lambda = \lambda/H_1) \quad i = -1, 1 \quad (12)$$

On en déduit que :

$$\left(\frac{\partial h(H/u_0)}{\partial \Lambda}\right)_{\Lambda=\lambda} = - \sum_{i \in \{-1,1\}} \left[-i \cdot P_0 \cdot P(\Lambda = \lambda/H_0) \log \frac{\alpha_i}{\alpha_i + \beta_i} - i \cdot (1 - P_0) \cdot P(\Lambda = \lambda/H_1) \log \frac{\beta_i}{\alpha_i + \beta_i} \right]$$

et puisque l'on a toujours $\frac{P(\Lambda = \lambda/H_1)}{P(\Lambda = \lambda/H_0)} = \lambda$, l'expression précédente peut s'écrire :

$$\left(\frac{\partial h(H/u_0)}{\partial \Lambda}\right)_{\Lambda=\lambda} = + \sum_{i \in \{-1,1\}} i \cdot P(\Lambda = \lambda/H_0) \cdot \left[P_0 \cdot \log \frac{\alpha_i}{\alpha_i + \beta_i} + \lambda \cdot (1 - P_0) \cdot \log \frac{\beta_i}{\alpha_i + \beta_i} \right] \quad (13)$$

La valeur du seuil λ qui minimise $h(H/u_0)$ est donc :

$$\lambda = - \frac{P_0}{1 - P_0} \cdot \frac{\sum_{i \in \{-1,1\}} i \cdot \log \frac{\alpha_i}{\alpha_i + \beta_i}}{\sum_{i \in \{-1,1\}} i \cdot \log \frac{\beta_i}{\alpha_i + \beta_i}} \quad (14)$$

En remplaçant α_i et β_i par leurs valeurs en fonction de P_0 , P_F et P_D (4), le seuil λ peut également s'écrire :

$$\lambda = - \frac{P_0}{1 - P_0} \cdot \frac{\log \frac{P_0 P_F}{P_0 P_F + (1 - P_0) P_D} - \log \frac{P_0(1 - P_F)}{P_0(1 - P_F) + (1 - P_0)(1 - P_D)}}{\log \frac{(1 - P_0)(1 - P_D)}{P_0(1 - P_F) + (1 - P_0)(1 - P_D)} - \log \frac{(1 - P_0) P_D}{P_0 P_F + (1 - P_0) P_D}} \quad (15)$$

Une formulation équivalente de ce seuil est proposée par P.K. Varshney [1], p.238.

Or, nous avons les quatre relations suivantes :

$$\begin{aligned} P(H_0/u_0 = 0) &= \frac{P(u_0 = 0/H_0) \cdot P_0}{P(u_0 = 0)} \\ &= \frac{(1 - P_F) P_0}{P(u_0 = 0/H_0) \cdot P_0 + P(u_0 = 0/H_1) \cdot (1 - P_0)} \\ &= \frac{P_0(1 - P_F)}{P_0(1 - P_F) + (1 - P_0)(1 - P_D)} \end{aligned}$$

$$P(H_0/u_0 = 1) = \frac{P_0 P_F}{P_0 P_F + (1 - P_0) P_D}$$

$$\begin{aligned} P(H_1/u_0 = 0) &= \frac{(1 - P_0)(1 - P_D)}{P_0(1 - P_F) + (1 - P_0)(1 - P_D)} \\ P(H_1/u_0 = 1) &= \frac{(1 - P_0) P_D}{P_0 P_F + (1 - P_0) P_D} \end{aligned} \quad (16)$$

De l'équation (15) et des relations (16), on déduit que λ s'exprime en fonction des probabilités *a posteriori* associées à chaque situation :

$$\lambda = \frac{P_0 \log P(H_0/u_0 = 1) - \log P(H_0/u_0 = 0)}{1 - P_0 \log P(H_1/u_0 = 0) - \log P(H_1/u_0 = 1)} \quad (17)$$

ou encore :

$$\lambda = \frac{P_0 \log 1/P(H_0/u_0 = 1) - \log 1/P(H_0/u_0 = 0)}{1 - P_0 \log 1/P(H_1/u_0 = 0) - \log 1/P(H_1/u_0 = 1)} \quad (18)$$

Chercher le détecteur qui minimise l'entropie conditionnelle $h(H/u_0)$ est donc équivalent à chercher le détecteur à rapport de vraisemblance $\Lambda(y) = \frac{p(y/H_1)}{p(y/H_0)}$ tel que :

$$\Lambda(y) \underset{u_0=0}{\overset{u_0=1}{\geq}} \frac{P_0}{(1 - P_0)} \cdot \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})} \quad (19)$$

où C_{ij} est un coût variable défini par $C_{ij} = \log \frac{1}{P(H_j/u_0 = i)}$ ($i, j = 0, 1$).

Remarques :

- Cette notation nous permet de surcroît une reformulation de l'entropie conditionnelle :

$$h(H/u_0) = E \left\{ \log \left(\frac{1}{P(H/u_0)} \right) \right\} = P_0 \cdot P_F \cdot (C_{10} - C_{00}) - (1 - P_0) \cdot P_D \cdot (C_{01} - C_{11}) + P_0 \cdot C_{00} + (1 - P_0) \cdot C_{01}$$

- Minimiser $h(H/u_0)$ à probabilité de fausse alarme fixée (voir figure 4) est équivalent à appliquer le critère de Neyman-Pearson.

Afin d'illustrer ces concepts, un exemple sera traité au paragraphe 5.

3. utilisation d'un critère entropique dans le cadre de la détection décentralisée parallèle

Dans cette partie, nous montrons que l'entropie peut être utilisée lors de l'optimisation d'une architecture de détection décentralisée parallèle. Dans un premier temps, nous proposons de considérer le problème de l'optimisation des différents opérateurs

locaux sans prendre en compte l'opérateur de fusion, que nous optimisons séparément dans un deuxième temps. L'intérêt principal de cette optimisation séparée est de proposer une solution simple à mettre en œuvre. Nous rappelons ensuite l'optimisation simultanée des détecteurs locaux et de l'opérateur de fusion.

3.1. optimisation des détecteurs locaux

Dans ce paragraphe, nous ne considérons que l'optimisation des détecteurs locaux sans prendre en compte le problème de la fusion. C'est une démarche similaire à celle développée dans [8] et [9] pour laquelle le critère d'optimisation utilisé était le critère de Bayes. Le système est composé de N détecteurs q_1, q_2, \dots, q_N associés à N capteurs Y_1, Y_2, \dots, Y_N (figure 5). Les mesures associées aux N capteurs sont toujours notées y_1, y_2, \dots, y_N . Les fonctions de densité de probabilités conditionnelles sous chaque hypothèse sont notées $p(y_1, y_2, \dots, y_N/H_k)$, $k = 0, 1$. Les décisions prises par chaque détecteur sont notées u_i ($i = 1, 2, \dots, N$) avec :

$$u_i = \begin{cases} 0, & \text{le détecteur « } i \text{ » décide que } H_0 \text{ est vraie} \\ 1, & \text{le détecteur « } i \text{ » décide que } H_1 \text{ est vraie} \end{cases} \quad (20)$$

Chaque décision binaire est caractérisée par ses probabilités de fausse alarme P_{F_i} et de détection P_{D_i} telles que : $P_{F_i} = P(u_i = 1/H_0)$ et $P_{D_i} = P(u_i = 1/H_1)$

Les décisions prises par les différents détecteurs sont supposées indépendantes.

Afin de simplifier le problème de l'optimisation des détecteurs locaux, considérons dans un premier temps le cas de 2 capteurs en parallèle ($N = 2$). Nous généralisons la démarche dans un deuxième temps.

a) Optimisation des détecteurs locaux pour $N = 2$ (2 capteurs en parallèle)

Les résultats obtenus dans le cas centralisé peuvent être repris :

$$\begin{aligned} h(H/u_1, u_2) &= - \sum_{i,j,k \in \{0,1\}} P(u_1 = i, u_2 = j, H_k) \\ &\quad \cdot \log P(H_k/u_1 = i, u_2 = j) \\ &= - \sum_{i,j,k \in \{0,1\}} P_k \cdot P(u_1 = i, u_2 = j/H_k) \\ &\quad \cdot \log \frac{P_k \cdot P(u_1 = i, u_2 = j/H_k)}{P(u_1 = i, u_2 = j)} \end{aligned} \quad (20)$$

et comme les décisions locales sont indépendantes entre elles,

$$\begin{aligned} h(H/u_1, u_2) &= \\ &= - \sum_{i,j,k \in \{0,1\}} P_k \cdot P(u_1 = i/H_k) \cdot P(u_2 = j/H_k) \cdot \log \\ &\quad \frac{P_k \cdot P(u_1 = i/H_k) \cdot P(u_2 = j/H_k)}{P(u_1 = i, u_2 = j/H_0) \cdot P_0 + P(u_1 = i, u_2 = j/H_1) \cdot (1 - P_0)} \end{aligned} \quad (21)$$

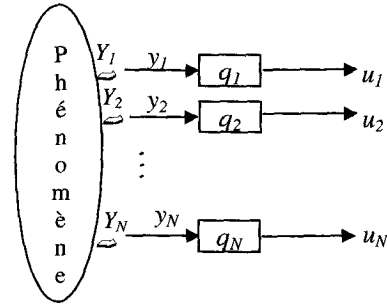


Figure 5. - N détecteurs en parallèle.

En développant suivant le terme k :

$$\begin{aligned} h(H/u_1, u_2) &= - \sum_{i,j \in \{0,1\}} P_0 \cdot P(u_1 = i/H_0) \cdot P(u_2 = j/H_0) \cdot \log \\ &\quad \frac{P_0 \cdot P(u_1 = i/H_0) \cdot P(u_2 = j/H_0)}{P(u_1 = i, u_2 = j/H_0) \cdot P_0 + P(u_1 = i, u_2 = j/H_1) \cdot (1 - P_0)} \\ &\quad - \sum_{i,j \in \{0,1\}} (1 - P_0) \cdot P(u_1 = i/H_1) \cdot P(u_2 = j/H_1) \cdot \log \\ &\quad \frac{(1 - P_0) \cdot P(u_1 = i/H_1) \cdot P(u_2 = j/H_1)}{P(u_1 = i, u_2 = j/H_0) \cdot P_0 + P(u_1 = i, u_2 = j/H_1) \cdot (1 - P_0)} \end{aligned}$$

et en utilisant les notations :

$$\alpha_{s_1 s_2} = P_0 \cdot P_{F_1}^{\frac{1+s_1}{2}} \cdot (1 - P_{F_1})^{\frac{1-s_1}{2}} \cdot P_{F_2}^{\frac{1+s_2}{2}} \cdot (1 - P_{F_2})^{\frac{1-s_2}{2}}$$

$$\beta_{s_1 s_2} = (1 - P_0) \cdot P_{D_1}^{\frac{1+s_1}{2}} \cdot (1 - P_{D_1})^{\frac{1-s_1}{2}} \cdot P_{D_2}^{\frac{1+s_2}{2}} \cdot (1 - P_{D_2})^{\frac{1-s_2}{2}}$$

$$s_1, s_2 \in \{-1, 1\}$$

nous obtenons l'équation :

$$\begin{aligned} h(H/u_1, u_2) &= - \sum_{s_1, s_2 \in \{-1, 1\}} \alpha_{s_1 s_2} \cdot \log \frac{\alpha_{s_1 s_2}}{\alpha_{s_1 s_2} + \beta_{s_1 s_2}} \\ &\quad + \beta_{s_1 s_2} \cdot \log \frac{\beta_{s_1 s_2}}{\alpha_{s_1 s_2} + \beta_{s_1 s_2}} \end{aligned} \quad (22)$$

Nous retrouvons les mêmes résultats que dans le cadre de la détection centralisée, à savoir :

$$P(\Lambda = \lambda_i/H_0) = - \left(\frac{\partial P_{F_i}}{\partial \Lambda} \right)_{\Lambda = \lambda_i}$$

et

$$P(\Lambda = \lambda_i/H_1) = - \left(\frac{\partial P_{D_i}}{\partial \Lambda} \right)_{\Lambda = \lambda_i} \quad i \in \{1, 2\}$$

$$\left(\frac{\partial \alpha_{s_1 s_2}}{\partial \Lambda} \right)_{\Lambda = \lambda_i} = -s_i \cdot P_0 \cdot P(\Lambda = \lambda_i/H_0)$$

et

$$\left(\frac{\partial \beta_{s_1 s_2}}{\partial \Lambda} \right)_{\Lambda=\lambda_i} = -s_i \cdot (1 - P_0) \cdot P(\Lambda = \lambda_i / H_1) \quad i \in \{1, 2\}$$

ce qui entraîne, pour $i \in \{1, 2\}$:

$$\left(\frac{\partial h(H/u_1 u_2)}{\partial \Lambda} \right)_{\Lambda=\lambda_i} = \sum_{s_1, s_2 \in \{-1, 1\}} [s_i \cdot P_0 \cdot P(\Lambda = \lambda_i / H_0) \cdot \log \frac{\alpha_{s_1 s_2}}{\alpha_{s_1 s_2} + \beta_{s_1 s_2}} + s_i \cdot (1 - P_0) \cdot \lambda_i \cdot P(\Lambda = \lambda_i / H_0) \cdot \log \frac{\beta_{s_1 s_2}}{\alpha_{s_1 s_2} + \beta_{s_1 s_2}}]$$

Les seuils λ_i qui minimisent $h(H/u_1, u_2)$ sont :

$$\lambda_i = -\frac{P_0}{1 - P_0} \cdot \frac{\sum_{s_1, s_2 \in \{-1, 1\}} s_i \cdot \log \frac{\alpha_{s_1 s_2}}{\alpha_{s_1 s_2} + \beta_{s_1 s_2}}}{\sum_{s_1, s_2 \in \{-1, 1\}} s_i \cdot \log \frac{\beta_{s_1 s_2}}{\alpha_{s_1 s_2} + \beta_{s_1 s_2}}} \quad i \in \{1, 2\} \quad (23)$$

b) Généralisation au cas de N détecteurs locaux en parallèle

Les résultats précédents peuvent être étendus au cas de N détecteurs en parallèle de la façon suivante.

Notons :

$$\alpha_{s_1, s_2, \dots, s_N} = P_0 \prod_{i \in \{1, 2, \dots, N\}} (P_{F_i})^{\frac{1+s_i}{2}} (1 - P_{F_i})^{\frac{1-s_i}{2}} \quad (24)$$

$$\beta_{s_1, s_2, \dots, s_N} = (1 - P_0) \prod_{i \in \{1, 2, \dots, N\}} (P_{D_i})^{\frac{1+s_i}{2}} (1 - P_{D_i})^{\frac{1-s_i}{2}} \quad (25)$$

$$s_1, s_2, \dots, s_N \in \{-1, 1\} :$$

L'entropie conditionnelle $h(H/u_1, u_2, \dots, u_N)$ peut alors s'écrire de la façon suivante :

$$h(H/u_1, u_2, \dots, u_N) = - \sum_{s_1, s_2, \dots, s_N \in \{-1, 1\}} \left[\alpha_{s_1, s_2, \dots, s_N} \cdot \log \frac{\alpha_{s_1, s_2, \dots, s_N}}{\alpha_{s_1, s_2, \dots, s_N} + \beta_{s_1, s_2, \dots, s_N}} + \beta_{s_1, s_2, \dots, s_N} \cdot \log \frac{\beta_{s_1, s_2, \dots, s_N}}{\alpha_{s_1, s_2, \dots, s_N} + \beta_{s_1, s_2, \dots, s_N}} \right] \quad (26)$$

Les détecteurs qui minimisent $h(H/u_1, u_2, \dots, u_N)$ consistent en un seuillage du rapport de vraisemblance et les seuils λ_i ($i = 1, 2, \dots, N$) associés à ces détecteurs sont déterminés à partir de la résolution d'un système de N équations non linéaires

couplées de la forme :

$$\lambda_i = -\frac{P_0}{1 - P_0} \cdot \frac{\sum_{s_1, s_2, \dots, s_N \in \{-1, 1\}} s_i \cdot \log \frac{\alpha_{s_1, s_2, \dots, s_N}}{\alpha_{s_1, s_2, \dots, s_N} + \beta_{s_1, s_2, \dots, s_N}}}{\sum_{s_1, s_2, \dots, s_N \in \{-1, 1\}} s_i \cdot \log \frac{\beta_{s_1, s_2, \dots, s_N}}{\alpha_{s_1, s_2, \dots, s_N} + \beta_{s_1, s_2, \dots, s_N}}} \quad i = \{1, 2, \dots, N\} \quad (27)$$

3.2. optimisation de l'opérateur de fusion

Les détecteurs locaux étant optimisés, considérons maintenant le problème de l'optimisation de l'opérateur de fusion (figure 6) à l'aide du critère entropique. Cette démarche est identique à celle proposée dans [10], [11], [12] avec le critère de Bayes.

Le système est composé de N détecteurs locaux q_i , associés à N capteurs Y_i , et prenant une décision locale $u_i = 0, 1$ ($i = 1, \dots, N$). Les décisions locales sont ensuite transmises à un opérateur de fusion qui les combine de façon à obtenir la décision finale u_0 .

$$u_0 = \begin{cases} 0, & \text{l'opérateur de fusion décide que } H_0 \text{ est vraie} \\ 1, & \text{l'opérateur de fusion décide que } H_1 \text{ est vraie} \end{cases}$$

Chaque décision binaire transmise à l'opérateur de fusion est caractérisée par ses probabilités de fausse alarme P_{F_i} et de détection P_{D_i} telles que :

$$P_{F_i} = P(u_i = 1/H_0) \text{ et } P_{D_i} = P(u_i = 1/H_1).$$

Les probabilités de fausse alarme et de détection du système complet sont notées P_F et P_D telles que :

$$P_F = P(u_0 = 1/H_0) \text{ et } P_D = P(u_0 = 1/H_1).$$

L'opérateur de fusion sera dès lors une fonction logique des N décisions binaires qui lui sont transmises. Dans le cas général, il y a 2^{2^N} opérateurs de fusion possibles. En pratique, il a été montré que la fonction de fusion doit être monotone [13] [12], ce qui

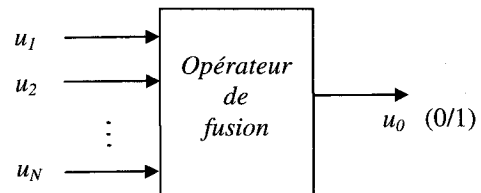


Figure 6. – L'opérateur de fusion.

réduit considérablement la taille de l'ensemble des fonctions de fusion admissibles (voir annexe 2).

Dans le cas particulier de deux capteurs, nous passons de 16 fonctions potentielles à 6 fonctions de fusion monotones (tableau 1).

Tableau 1. – Les fonctions de fusion monotones dans le cas de deux détecteurs en parallèle.

u_1	u_2	f_1	f_2	f_4	f_6	f_8	f_{16}
		et			ou		
0	0	0	0	0	0	0	1
0	1	0	0	0	1	1	1
1	0	0	0	1	0	1	1
1	1	0	1	1	1	1	1

Au paragraphe 2.1, nous avons démontré que $h(H/u_0)$ était une fonction de P_0 , P_D et P_F . Le problème consiste alors à déterminer la règle de fusion monotone qui minimise l'entropie conditionnelle $h(H/u_0)$.

Dans le cas où les décisions locales sont indépendantes, les probabilités de fausse alarme P_F et de détection P_D du système complet peuvent s'exprimer en fonction des probabilités P_{F_i} et P_{D_i} ($i = 1, 2$) des différents détecteurs (tableau 2).

Tableau 2. – Valeurs de P_F et P_D suivant la fonction de fusion.

	f_1	f_2	f_4	f_6	f_8	f_{16}
P_D	0	$P_{D_1}P_{D_2}$	P_{D_1}	P_{D_2}	$P_{D_1} + P_{D_2} - P_{D_1}P_{D_2}$	1
P_F	0	$P_{F_1}P_{F_2}$	P_{F_1}	P_{F_2}	$P_{F_1} + P_{F_2} - P_{F_1}P_{F_2}$	1

Il suffit alors de remplacer, dans l'équation (26), P_F et P_D par leurs valeurs en fonction de P_{F_i} et P_{D_i} pour toutes les règles de fusion possibles; et la règle de fusion optimale sera celle qui minimise $h(H/u_0)$.

3.3. optimisation simultanée des détecteurs locaux et de l'opérateur de fusion

Considérons l'optimisation simultanée des détecteurs locaux et de l'opérateur de fusion (figure 7). L'optimisation de ces deux composantes de façon simultanée, tant du point de vue Bayésien [14] [15] [16] que du point de vue de Neyman-Pearson [17] [18], a

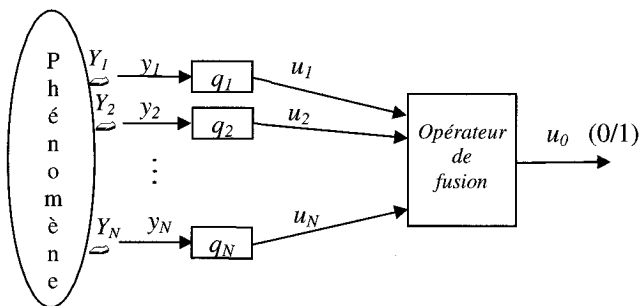


Figure 7. – Architecture de détection décentralisée parallèle.

été largement étudiée dans la littérature. Une démarche identique est développée ici en utilisant un critère entropique [1].

3.3.1. expression de l'entropie $h(H/u_0)$

Nous cherchons à optimiser cette architecture en minimisant l'entropie conditionnelle $h(H/u_0)$. Or, nous avons vu au paragraphe 2.1, que l'entropie pouvait se mettre sous la forme suivante :

$$h(H/u_0) = E \left\{ \log \left(\frac{1}{P(H/u_0)} \right) \right\} = - \sum_{i \in \{-1,1\}} \left[\alpha_i \log \frac{\alpha_i}{\alpha_i + \beta_i} + \beta_i \log \frac{\beta_i}{\alpha_i + \beta_i} \right] \quad (28)$$

avec :

$$\alpha_1 = P_0 P_F \quad \beta_1 = (1 - P_0) P_D \\ \alpha_{-1} = P_0 (1 - P_F) \quad \text{et} \quad \beta_{-1} = (1 - P_0) (1 - P_D) \quad (29)$$

Minimiser $h(H/u_0)$ revient alors à minimiser une fonction « risque moyen » pour laquelle les fonctions de coût ne sont pas constantes mais dépendent des probabilités a posteriori. L'optimisation du système se fait « élément par élément ». Dans un premier temps, nous cherchons à optimiser l'opérateur de fusion, les détecteurs locaux étant supposés fixés.

3.3.2. optimisation de l'opérateur de fusion

En utilisant la propriété (7), les relations (28) et (29) entraînent :

$$\frac{\partial h(H/u_0)}{\partial v} = - \sum_{i \in \{-1,1\}} \left[\frac{\partial \alpha_i}{\partial v} \log \frac{\alpha_i}{\alpha_i + \beta_i} + \frac{\partial \beta_i}{\partial v} \log \frac{\beta_i}{\alpha_i + \beta_i} \right] \quad (30)$$

pour une variable quelconque v . Ce qui nous amène à la relation suivante :

$$\frac{\partial h(H/u_0)}{\partial v} = - \sum_{i \in \{-1,1\}} \left[i \cdot P_0 \cdot \frac{\partial P_F}{\partial v} \log \frac{\alpha_i}{\alpha_i + \beta_i} + i \cdot (1 - P_0) \cdot \frac{\partial P_D}{\partial v} \log \frac{\beta_i}{\alpha_i + \beta_i} \right] \quad (31)$$

Et compte tenu que :

$$C_{10} = \log \frac{1}{P(H_0/u_0 = 1)} \quad (32) \\ = \log \frac{P(u_0 = 1/H_0)P_0 + P(u_0 = 1/H_1)(1 - P_0)}{P(u_0 = 1/H_0)P_0} \\ = \log \frac{P_0 P_F + (1 - P_0) P_D}{P_0 P_F} = \log \frac{\alpha_1 + \beta_1}{\alpha_1} \\ C_{00} = \log \frac{P_0(1 - P_F) + (1 - P_0)(1 - P_D)}{P_0(1 - P_F)} \\ = \log \frac{\alpha_{-1} + \beta_{-1}}{\alpha_{-1}}$$

$$C_{11} = \log \frac{P_0 P_F + (1 - P_0) P_D}{(1 - P_0) P_D} = \log \frac{\alpha_1 + \beta_1}{\beta_1}$$

$$C_{01} = \log \frac{P_0(1 - P_F) + (1 - P_0)(1 - P_D)}{(1 - P_0)(1 - P_D)}$$

$$= \log \frac{\alpha_{-1} + \beta_{-1}}{\beta_{-1}}$$

nous obtenons la relation :

$$\frac{\partial h(H/u_0)}{\partial v} = P_0(C_{10} - C_{00}) \frac{\partial P_F}{\partial v} - (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_D}{\partial v} \quad (33)$$

Le signe de cette expression nous permet de déterminer le sens de variation de $h(H/u_0)$ en fonction de v .

D'autre part, si l'on appelle $u = (u_1, \dots, u_N)$ le vecteur des décisions prises par les détecteurs locaux, on a :

$$P_D = P(u_0 = 1/H_1) = \sum_u P(u_0 = 1/u) \cdot P(u/H_1) \quad (34)$$

$$P_F = P(u_0 = 1/H_0) = \sum_u P(u_0 = 1/u) \cdot P(u/H_0)$$

Et si l'on pose $v = P(u_0 = 1/u^*)$, la probabilité de décider $u_0 = 1$ sachant que le vecteur u est égal à la solution particulière $u^* = (u_1^*, \dots, u_N^*)$, l'optimisation « élément par élément » donne :

$$\frac{\partial h(H/u_0)}{\partial P(u_0 = 1/u^*)} = P_0(C_{10} - C_{00}) \frac{\partial P_F}{\partial P(u_0 = 1/u^*)} - (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_D}{\partial P(u_0 = 1/u^*)} \quad (35)$$

$$\frac{\partial h(H/u_0)}{\partial P(u_0 = 1/u^*)} = P_0(C_{10} - C_{00}) P(u^*/H_0) - (1 - P_0)(C_{01} - C_{11}) P(u^*/H_1) \quad (36)$$

L'opérateur de fusion est obtenu en remarquant que, lorsque la quantité (36) est négative (*resp.* positive), l'entropie $h(H/u_0)$ est une fonction décroissante (*resp.* croissante) de la variable $P(u_0 = 1/u^*)$. Par conséquent, afin de minimiser $h(H/u_0)$, il suffit de fixer $P(u_0 = 1/u^*) = 1$ lorsque $h(H/u_0)$ est une fonction décroissante de $P(u_0 = 1/u^*)$, et $P(u_0 = 1/u^*) = 0$ lorsque $h(H/u_0)$ est une fonction croissante de $P(u_0 = 1/u^*)$. Ce qui équivaut à appliquer la règle suivante :

$$P_0(C_{10} - C_{00}) P(u^*/H_0) - (1 - P_0)(C_{01} - C_{11}) P(u^*/H_1) \underset{P(u_0=1/u^*)=1}{\underset{P(u_0=1/u^*)=0}{\geq}} 0 \quad (37)$$

De plus, en faisant l'hypothèse raisonnable que le coût d'une mauvaise décision est supérieur au coût d'une bonne décision, c'est-à-dire que $C_{10} > C_{00}$ et $C_{01} > C_{11}$, la règle (37) peut se mettre sous la forme suivante :

$$\frac{P(u^*/H_1)}{P(u^*/H_0)} \underset{u_0=0}{\underset{u_0=1}{\geq}} \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})} \quad (38)$$

La règle (38) est similaire à la règle obtenue en appliquant les critères de Bayes [15] [4] [16] ou de Neyman-Pearson [17]. Les fonctions de coût ne sont pas des constantes mais dépendent des probabilités *a posteriori*.

3.3.3. optimisation des détecteurs locaux

L'optimisation « élément par élément » du détecteur k ($k = 1, \dots, N$) peut être obtenue en dérivant $h(H/u_0)$ par rapport à P_{F_k} . Pour cela, il suffit de poser $v = P_{F_k}$ dans l'équation (33), et on obtient :

$$\frac{\partial h(H/u_0)}{\partial P_{F_k}} = P_0(C_{10} - C_{00}) \frac{\partial P_F}{\partial P_{F_k}} - (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_D}{\partial P_{F_k}} \quad (39)$$

Sachant que le seuil λ_k associé au détecteur k est donné par $\lambda_k = \frac{\partial P_{D_k}}{\partial P_{F_k}}$ (voir annexe 3), on a :

$$\frac{\partial P_D}{\partial P_{F_k}} = \frac{\partial P_D}{\partial P_{D_k}} \lambda_k \quad (40)$$

et en substituant (40) dans (39), on obtient :

$$\frac{\partial h(H/u_0)}{\partial P_{F_k}} = P_0(C_{10} - C_{00}) \frac{\partial P_F}{\partial P_{F_k}} - (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_D}{\partial P_{D_k}} \lambda_k \quad (41)$$

Minimiser $h(H/u_0)$ revient alors à annuler (41), ce qui revient à choisir :

$$\lambda_k = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) \frac{\partial P_F}{\partial P_{F_k}}}{(C_{01} - C_{11}) \frac{\partial P_D}{\partial P_{D_k}}} \quad (k = 1, \dots, N) \quad (42)$$

Une solution commune aux équations (38) et (42) est une solution de l'optimisation « élément par élément » utilisant le critère entropique du problème de détection décentralisée de la figure 7.

3.3.4. cas particulier de deux détecteurs en parallèle

En considérant deux capteurs dont les observations sont indépendantes, une expression explicite des seuils peut être obtenue en fonction de l'opérateur de fusion qui a été retenu. Ainsi, l'équation (42) permet d'écrire :

$$\lambda_1 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) \frac{\partial P_F}{\partial P_{F_1}}}{(C_{01} - C_{11}) \frac{\partial P_D}{\partial P_{D_1}}}$$

et

$$\lambda_2 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) \frac{\partial P_F}{\partial P_{F_2}}}{(C_{01} - C_{11}) \frac{\partial P_D}{\partial P_{D_2}}} \quad (43)$$

- Dans le cas particulier où l'opérateur de fusion est un « et logique », les expressions des probabilités de fausse alarme P_F et de détection P_D du système complet peuvent être calculées par : $P_D = P_{D_1} \cdot P_{D_2}$ et $P_F = P_{F_1} \cdot P_{F_2}$

On en déduit alors :

$$\lambda_1 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) P_{F_2}}{(C_{01} - C_{11}) P_{D_2}}$$

et

$$\lambda_2 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) P_{F_1}}{(C_{01} - C_{11}) P_{D_1}} \quad (44)$$

- Dans le cas particulier où l'opérateur de fusion est un « ou logique », les expressions des probabilités de fausse alarme P_F et de détection P_D du système complet sont données par : $P_D = P_{D_1} + P_{D_2} - P_{D_1} \cdot P_{D_2}$ et $P_F = P_{F_1} + P_{F_2} - P_{F_1} \cdot P_{F_2}$

On en déduit alors :

$$\lambda_1 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00})(1 - P_{F_2})}{(C_{01} - C_{11})(1 - P_{D_2})}$$

et

$$\lambda_2 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00})(1 - P_{F_1})}{(C_{01} - C_{11})(1 - P_{D_1})} \quad (45)$$

Les équations (44) (resp. (45)) sont couplées. Leur résolution permet de trouver les valeurs des seuils λ_1 et λ_2 qui minimisent $h(H/u_0)$ dans le cas où l'opérateur de fusion est un « et logique » (resp. un « ou logique »).

4. utilisation d'un critère entropique dans le cadre de la détection décentralisée série

Dans cette partie, nous considérons le problème de la détection décentralisée série de la figure 8. Chaque détecteur local reçoit une

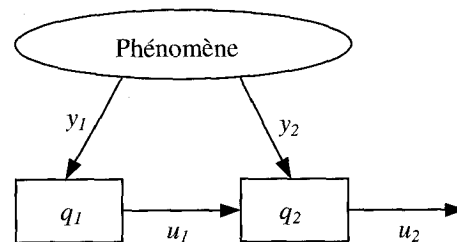


Figure 8. – Architecture de détection décentralisée série à deux détecteurs.

information issue d'un capteur et transmet un message binaire à son successeur. La décision du premier détecteur est basée sur les informations issues d'un seul capteur et c'est le dernier détecteur qui élabore la décision finale [19] [20] [16].

Nous étudions un système en tandem formé de deux détecteurs q_1 et q_2 . Les mesures issues de chaque capteur sont notées respectivement y_1 et y_2 . Les fonctions de densité de probabilités conditionnelles sous chaque hypothèse sont notées $p(y_1, y_2/H_k)$, $k = 0, 1$. Le détecteur q_1 élabore sa décision u_1 à partir de y_1 . Cette décision est ensuite transmise au détecteur q_2 qui élabore la décision finale u_2 à partir de u_1 et de y_2 .

Chaque décision binaire est telle que :

$$u_i = \begin{cases} 0, & \text{le détecteur « i » décide que } H_0 \text{ est vraie} \\ 1, & \text{le détecteur « i » décide que } H_1 \text{ est vraie} \end{cases}$$

Chaque détecteur local est caractérisé par ses probabilités de fausse alarme P_{F_i} et de détection P_{D_i} .

4.1. expression de l'entropie conditionnelle

Notre approche consiste à optimiser l'architecture de ce système en minimisant l'entropie conditionnelle $h(H/u_2)$. Or, nous avons vu aux paragraphes 2.1 et 3.3.1, que l'entropie pouvait se mettre sous la forme suivante :

$$\begin{aligned} h(H/u_2) &= E \left\{ \log \left(\frac{1}{P(H/u_2)} \right) \right\} \\ &= - \sum_{i \in \{-1, 1\}} \left[\alpha_i \log \frac{\alpha_i}{\alpha_i + \beta_i} + \beta_i \log \frac{\beta_i}{\alpha_i + \beta_i} \right] \quad (46) \end{aligned}$$

avec :

$$\begin{aligned} \alpha_1 &= P_0 P_{F_2} & \beta_1 &= (1 - P_0) P_{D_2} \\ \alpha_{-1} &= P_0 (1 - P_{F_2}) & \beta_{-1} &= (1 - P_0) (1 - P_{D_2}) \end{aligned} \quad (47)$$

Minimiser $h(H/u_2)$ revient alors à minimiser une fonction « risque moyen » pour laquelle les fonctions de coût ne sont pas constantes, mais dépendent des probabilités *a posteriori*. L'optimisation de ce système se fait « élément par élément ». Dans un premier temps, le deuxième détecteur est optimisé, le premier détecteur étant supposé fixé. De même, lors de l'optimisation du premier détecteur, le second détecteur sera supposé fixé.

4.2. optimisation du second détecteur

Comme au paragraphe 3.3.2, dès lors que nous fixons les coûts :

$$C_{10} = \log \frac{P_0 P_{F_2} + (1 - P_0) P_{D_2}}{P_0 P_{F_2}} \quad (48)$$

$$C_{00} = \log \frac{P_0(1 - P_{F_2}) + (1 - P_0)(1 - P_{D_2})}{P_0(1 - P_{F_2})}$$

$$C_{11} = \log \frac{P_0 P_{F_2} + (1 - P_0) P_{D_2}}{(1 - P_0) P_{D_2}}$$

$$C_{01} = \log \frac{P_0(1 - P_{F_2}) + (1 - P_0)(1 - P_{D_2})}{(1 - P_0)(1 - P_{D_2})}$$

Nous aboutissons à la relation :

$$\frac{\partial h(H/u_2)}{\partial v} = P_0(C_{10} - C_{00}) \frac{\partial P_{F_2}}{\partial v} - (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_{D_2}}{\partial v} \quad (49)$$

Le signe de cette expression nous permet de déterminer le sens de variation de $h(H/u_2)$ en fonction de la variable v .

D'autre part, on a :

$$\begin{aligned} P_{D_2} &= P(u_2 = 1/H_1) = \\ &\sum_{u_1, y_2} P(u_2 = 1/u_1, y_2) \cdot P(u_1, y_2/H_1) \\ P_{F_2} &= P(u_2 = 1/H_0) = \\ &\sum_{u_1, y_2} P(u_2 = 1/u_1, y_2) \cdot P(u_1, y_2/H_0) \end{aligned} \quad (50)$$

Et si l'on pose $v = P(u_2 = 1/u_1^*, y_2^*)$, la probabilité de décider $u_2 = 1$ sachant que u_1 et y_2 prennent les valeurs particulières u_1^* et y_2^* , l'optimisation « élément par élément » donne :

$$\begin{aligned} \frac{\partial h(H/u_2)}{\partial P(u_2 = 1/u_1^*, y_2^*)} &= P_0(C_{10} - C_{00}) \frac{\partial P_{F_2}}{\partial P(u_2 = 1/u_1^*, y_2^*)} \\ &- (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_{D_2}}{\partial P(u_2 = 1/u_1^*, y_2^*)} \end{aligned} \quad (51)$$

$$\begin{aligned} \frac{\partial h(H/u_2)}{\partial P(u_2 = 1/u_1^*, y_2^*)} &= P_0(C_{10} - C_{00}) P(u_1^*, y_2^*/H_0) \\ &- (1 - P_0)(C_{01} - C_{11}) P(u_1^*, y_2^*/H_1) \end{aligned} \quad (52)$$

L'opérateur de fusion est obtenu en observant que, lorsque la quantité (52) est négative (resp. positive), l'entropie $h(H/u_2)$ est une fonction décroissante (resp. croissante) de la variable $P(u_2 = 1/u_1^*, y_2^*)$. Par conséquent, minimiser $h(H/u_2)$ revient à fixer $P(u_2 = 1/u_1^*, y_2^*) = 1$ lorsque $h(H/u_2)$ est une fonction décroissante de $P(u_2 = 1/u_1^*, y_2^*)$, et $P(u_2 = 1/u_1^*, y_2^*) = 0$ lorsque $h(H/u_2)$ est une fonction croissante de $P(u_2 = 1/u_1^*, y_2^*)$. Ce qui équivaut à utiliser la règle suivante :

$$\begin{aligned} P_0(C_{10} - C_{00}) P(u_1^*, y_2^*/H_0) \\ - (1 - P_0)(C_{01} - C_{11}) P(u_1^*, y_2^*/H_1) \begin{cases} P(u_2=1/u_1^*, y_2^*)=0 \\ \geq \\ P(u_2=1/u_1^*, y_2^*)=1 \end{cases} 0 \end{aligned} \quad (53)$$

En faisant l'hypothèse raisonnable que le coût d'une mauvaise décision est supérieur au coût d'une bonne décision, c'est-à-dire que $C_{10} > C_{00}$ et $C_{01} > C_{11}$, la règle (53) peut se mettre sous la forme :

$$\frac{P(u_1^*, y_2^*/H_1)}{P(u_1^*, y_2^*/H_0)} \underset{u_2=0}{\overset{u_2=1}{\geq}} \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00})}{(C_{01} - C_{11})} \quad (54)$$

De plus, si u_1 et y_2 sont indépendants, on a :

$$\frac{P(u_1^*, y_2^*/H_1)}{P(u_1^*, y_2^*/H_0)} = \frac{P(u_1^*/H_1) P(y_2^*/H_1)}{P(u_1^*/H_0) P(y_2^*/H_0)} \quad (55)$$

et la règle (54) peut s'écrire :

$$\frac{P(y_2^*/H_1)}{P(y_2^*/H_0)} \underset{u_2=0}{\overset{u_2=1}{\geq}} \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) P(u_1^*/H_0)}{(C_{01} - C_{11}) P(u_1^*/H_1)} \quad (56)$$

La règle (56) est similaire à celle obtenue en utilisant le critère de Bayes [20] [16], à la différence que les fonctions de coût ne sont pas des constantes mais dépendent des probabilités *a posteriori*. De plus, il faut remarquer que le seuil de droite correspond en fait à deux seuils différents, l'un lorsque $u_1^* = 0$, et l'autre lorsque $u_1^* = 1$. Ces deux seuils sont notés λ_2^1 et λ_2^0 , et dépendent des probabilités de fausse alarme P_{F_1} et de détection P_{D_1} du détecteur q_1 tels que :

$$\begin{aligned} \lambda_2^1 &= \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) P(u_1 = 1/H_0)}{(C_{01} - C_{11}) P(u_1 = 1/H_1)} \\ &= \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) P_{F_1}}{(C_{01} - C_{11}) P_{D_1}} \end{aligned} \quad (57)$$

$$\begin{aligned} \lambda_2^0 &= \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) P(u_1 = 0/H_0)}{(C_{01} - C_{11}) P(u_1 = 0/H_1)} \\ &= \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) (1 - P_{F_1})}{(C_{01} - C_{11}) (1 - P_{D_1})} \end{aligned} \quad (58)$$

4.3. optimisation du premier détecteur

L'optimisation du premier détecteur peut être obtenue en dérivant $h(H/u_2)$ par rapport à P_{F_1} . En posant $v = P_{F_1}$ dans l'équation (49), on obtient :

$$\frac{\partial h(H/u_2)}{\partial P_{F_1}} = P_0(C_{10} - C_{00}) \frac{\partial P_{F_2}}{\partial P_{F_1}} - (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_{D_2}}{\partial P_{F_1}} \quad (59)$$

De plus, sachant que le seuil λ_1 au niveau du premier détecteur est donné par : $\lambda_1 = \frac{\partial P_{D_1}}{\partial P_{F_1}}$ et que l'on a :

$$\frac{\partial P_{D_2}}{\partial P_{F_1}} = \frac{\partial P_{D_2}}{\partial P_{D_1}} \lambda_1 \quad (60)$$

en substituant (60) dans (59), on obtient la relation suivante :

$$\frac{\partial h(H/u_2)}{\partial P_{F_1}} = P_0(C_{10} - C_{00}) \frac{\partial P_{F_2}}{\partial P_{F_1}} - (1 - P_0)(C_{01} - C_{11}) \frac{\partial P_{D_2}}{\partial P_{D_1}} \lambda_1 \quad (61)$$

Afin de minimiser $h(H/u_2)$, il suffit que l'expression (61) soit nulle, ce qui revient à choisir λ_1 de la façon suivante :

$$\lambda_1 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00}) \frac{\partial P_{F_2}}{\partial P_{F_1}}}{(C_{01} - C_{11}) \frac{\partial P_{D_2}}{\partial P_{D_1}}} \quad (62)$$

De plus :

$$\begin{aligned} P_{F_2} &= P(u_2 = 1/H_0) \\ P_{F_2} &= P(u_2 = 1/u_1 = 0, H_0) \cdot P(u_1 = 0/H_0) \\ &\quad + P(u_2 = 1/u_1 = 1, H_0) \cdot P(u_1 = 1/H_0) \quad (63) \\ P_{F_2} &= P_{F_2}(\lambda_2^0) \cdot (1 - P_{F_1}) + P_{F_2}(\lambda_2^1) \cdot P_{F_1} \end{aligned}$$

De même :

$$P_{D_2} = P_{D_2}(\lambda_2^0) \cdot (1 - P_{D_1}) + P_{D_2}(\lambda_2^1) \cdot P_{D_1} \quad (64)$$

où $P_{F_2}(\lambda_2^j)$ et $P_{D_2}(\lambda_2^j)$ ($j = 0, 1$) représentent les probabilités de fausse alarme et de détection du second détecteur calculées en utilisant λ_2^j comme seuil. Sachant que le second détecteur est supposé fixé, on en déduit que :

$$\frac{\partial P_{F_2}}{\partial P_{F_1}} = P_{F_2}(\lambda_2^1) - P_{F_2}(\lambda_2^0) \quad \text{et} \quad \frac{\partial P_{D_2}}{\partial P_{D_1}} = P_{D_2}(\lambda_2^1) - P_{D_2}(\lambda_2^0) \quad (65)$$

Par conséquent, le seuil λ_1 peut s'écrire de la façon suivante :

$$\lambda_1 = \frac{P_0}{1 - P_0} \frac{(C_{10} - C_{00})(P_{F_2}(\lambda_2^1) - P_{F_2}(\lambda_2^0))}{(C_{01} - C_{11})(P_{D_2}(\lambda_2^1) - P_{D_2}(\lambda_2^0))} \quad (66)$$

L'équation (66) est similaire à celle obtenue en appliquant le critère de Bayes [20] [16]. Optimiser le système revient donc à résoudre 3 équations non linéaires couplées de la forme (66), (57) et (58). Une solution commune à ces 3 équations est une solution de l'optimisation « élément par élément » utilisant le critère entropique du problème de détection décentralisée série de la figure 8.

En pratique, connaissant les lois de densités de probabilités conditionnelles sous chaque hypothèse pour les deux détecteurs, il suffit de fixer le seuil λ_1 pour le premier détecteur ainsi que les deux seuils λ_2^0 et λ_2^1 pour le second détecteur afin de calculer respectivement les couples (P_{F_1}, P_{D_1}) , $(P_{F_2}^0, P_{D_2}^0)$ et $(P_{F_2}^1, P_{D_2}^1)$. La connaissance de ces trois couples détermine les probabilités P_{F_2} et P_{D_2} pour le second détecteur, ce qui permet le calcul de la fonctionnelle $h(H/u_2)$ (46). Les trois seuils optimaux sont alors ceux pour lesquels la fonctionnelle $h(H/u_2)$ est minimale.

5. exemple illustratif

Considérons un système formé de deux capteurs Y_1 et Y_2 dont les observations y_1 et y_2 sont indépendantes. Les fonctions de densité de probabilités sous chaque hypothèse sont des gaussiennes. Sous l'hypothèse H_0 , ces fonctions de densité de probabilités sont supposées être de moyenne 0 et de variance 1. Sous l'hypothèse H_1 , ces fonctions sont supposées être de variance 1, et respectivement de moyenne 1 pour Y_1 et 1.5 pour Y_2 . De façon intuitive, il paraît évident que le capteur Y_2 permettra une meilleure détection, le chevauchement des deux gaussiennes étant moins important pour ce capteur. Les résultats obtenus pour chaque architecture (paragraphes 5.1 à 5.3) en appliquant le critère de Bayes et le critère entropique sont discutés ensemble au paragraphe 5.4. Dans cette optique, nous disposons de deux outils de comparaison : les courbes C.O.R. (Caractéristique Opérationnelle du Récepteur) et les courbes de probabilités d'erreur.

5.1. détection centralisée

Que l'on utilise le critère de Bayes ou le critère entropique, la règle de décision optimale pour chaque capteur est, d'après (19) :

$$\Lambda(y_i) \underset{u_0=0}{\overset{u_0=1}{\geq}} \lambda_i \quad i = 1, 2$$

Nous avons tracé la probabilité d'erreur en fonction de P_0 ainsi que la courbe COR en prenant P_0 comme paramètre pour chacun des deux capteurs et en appliquant les deux critères : les figures 9 et 10 représentent respectivement les courbes COR et les probabilités d'erreur obtenues en utilisant le critère de Bayes avec $C_{00} = C_{11} = 0$ et $C_{01} = C_{10} = 1$, où C_{ij} est le coût correspondant à la décision $u_0 = i$ lorsque H_j est vraie.

Les figures 11 et 12 représentent respectivement les courbes COR et les probabilités d'erreur obtenues en minimisant $h(H/u_0)$.

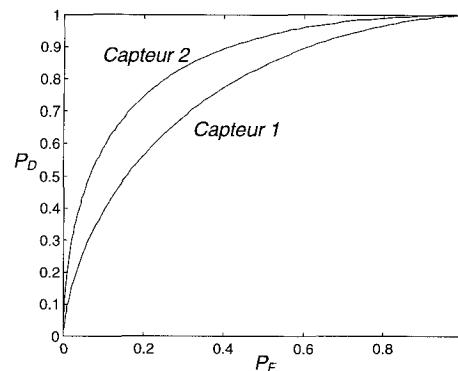


Figure 9. – Courbes COR obtenues en utilisant le critère de Bayes en détection centralisée.

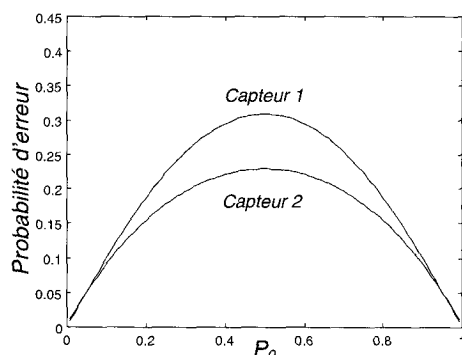


Figure 10. – Courbes de probabilité d'erreur obtenues en utilisant le critère de Bayes en détection centralisée.

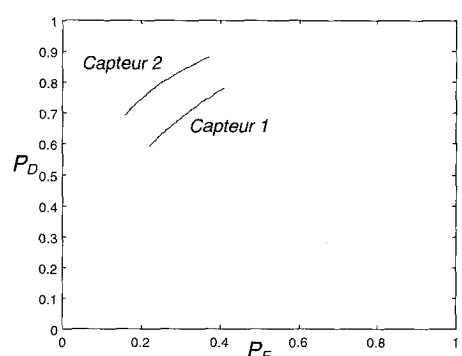


Figure 11. – Courbes COR obtenues en utilisant le critère entropique en détection centralisée.

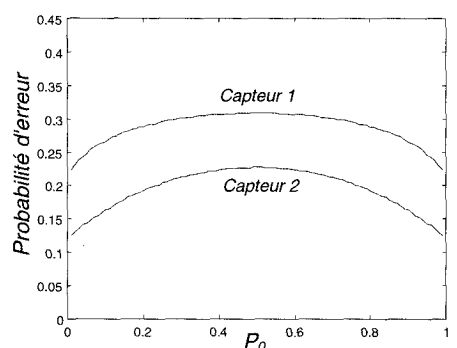


Figure 12. – Courbes de probabilité d'erreur obtenues en utilisant le critère entropique en détection centralisée.

5.2. détection décentralisée parallèle

5.2.1. optimisation en deux étapes

Dans le cas Bayésien, les coûts associés aux différentes situations sont notés C_{ijk} , $i, j, k = 0, 1$ où C_{ijk} représente le coût de décider $u_1 = i$ pour le premier détecteur, $u_2 = j$ pour le second détecteur alors que H_k est vraie. Afin d'optimiser les détecteurs locaux, on minimise l'espérance de ce coût $E\{C_{ijk}\}$. Puis, les détecteurs

locaux étant déterminés, on optimise l'opérateur de fusion. On appelle C_{ij} le coût correspondant à la décision $u_0 = i$ lorsque H_j est vraie et on minimise $E(C_{ij})$. Dans cet exemple, nous avons pris $C_{000} = C_{111} = 0$, $C_{010} = C_{100} = C_{011} = C_{101} = 1$, et $C_{110} = C_{001} = 2$ pour l'optimisation des détecteurs locaux et $C_{00} = C_{11} = 0$ et $C_{01} = C_{10} = 1$ pour l'optimisation de l'opérateur de fusion.

Dans le cas entropique, nous avons optimisé les détecteurs locaux en minimisant l'entropie conditionnelle $h(H/u_1, u_2)$. Puis, les détecteurs locaux étant déterminés, nous avons optimisé l'opérateur de fusion en minimisant $h(H/u_0)$.

Que l'on utilise le critère de Bayes ou le critère entropique, la règle de décision optimale au niveau de chaque détecteur est, d'après

$$(23) : \quad \Lambda(y_i) \underset{u_i=0}{\overset{u_i=1}{\geq}} \lambda_i \quad i = 1, 2$$

Nous avons tracé la probabilité d'erreur en fonction de P_0 ainsi que la courbe COR pour chacune des fonctions de fusion possibles (tableau 1) en utilisant les deux critères.

Les figures 13 et 14 représentent respectivement les courbes COR et les probabilités d'erreur obtenues en utilisant le critère de Bayes.

Les figures 15 et 16 représentent les courbes COR et les probabilités d'erreur obtenues en minimisant $h(H/u_0)$.

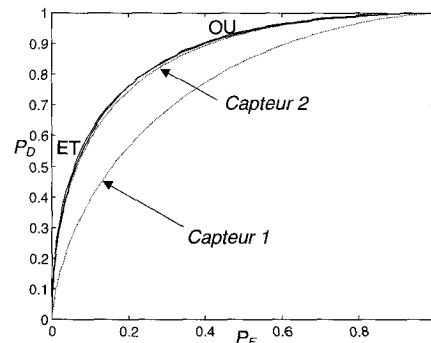


Figure 13. – Courbes COR obtenues en utilisant le critère de Bayes en détection décentralisée parallèle - optimisation en 2 étapes.

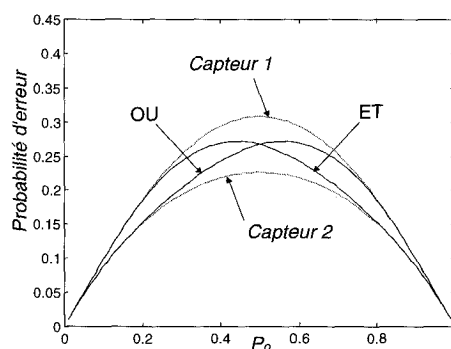


Figure 14. – Courbes de probabilité d'erreur obtenues en utilisant le critère de Bayes en détection décentralisée parallèle - optimisation en 2 étapes.

Optimisation entropique des systèmes de détection distribuée

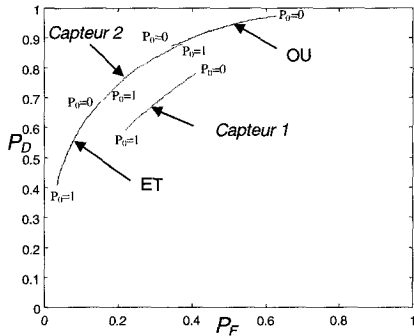


Figure 15. – Courbes COR obtenues en utilisant le critère entropique en détection décentralisée parallèle - optimisation en 2 étapes.

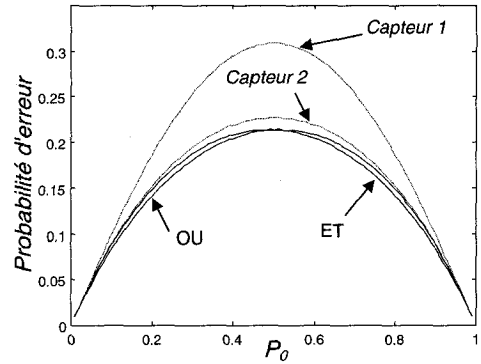


Figure 18. – Courbes de probabilité d'erreur obtenues en utilisant le critère de Bayes en détection décentralisée parallèle - optimisation simultanée.

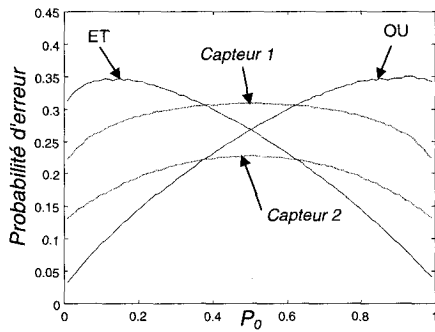


Figure 16. – Courbes de probabilité d'erreur obtenues en utilisant le critère entropique en détection décentralisée parallèle - optimisation en 2 étapes.

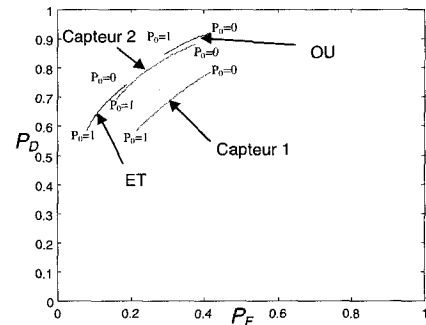


Figure 19. – Courbes COR obtenues en utilisant le critère entropique en détection décentralisée parallèle - optimisation simultanée.

5.2.2. Optimisation simultanée

Le critère de Bayes consiste à affecter un coût C_{ij} correspondant à la décision $u_0 = i$ lorsque H_j est vraie et à minimiser $E(C_{ij})$. Le critère entropique consiste à minimiser $h(H/u_0)$. Les figures 17 et 18 représentent respectivement les courbes COR et les probabilités d'erreur obtenues en utilisant le critère de Bayes pour lequel nous avons considéré les coûts $C_{00} = C_{11} = 0$ et $C_{01} = C_{10} = 1$.

Les figures 19 et 20 représentent respectivement les courbes COR et les probabilités d'erreur obtenues en minimisant $h(H/u_0)$.

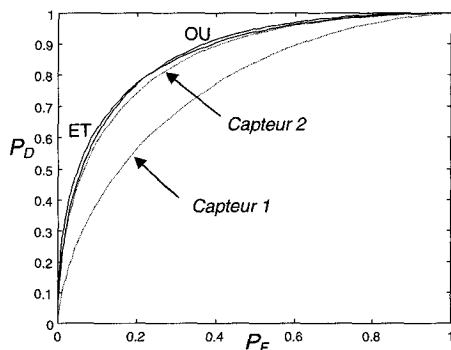


Figure 17. – Courbes COR obtenues en utilisant le critère de Bayes en détection décentralisée parallèle - optimisation simultanée.

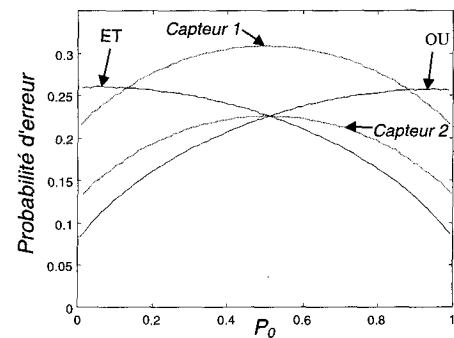


Figure 20. – Courbes de probabilité d'erreur obtenues en utilisant le critère entropique en détection décentralisée parallèle - optimisation simultanée.

5.3. détection décentralisée série

Le critère de Bayes consiste à affecter un coût C_{ij} correspondant à la décision $u_2 = i$ lorsque H_j est vraie, et à minimiser $E(C_{ij})$. Le critère entropique consiste à minimiser $h(H/u_2)$.

Les Figures 21 et 22 représentent respectivement les courbes COR et les probabilités d'erreur obtenues en utilisant le critère de Bayes pour lequel nous avons considéré $C_{00} = C_{11} = 0$ et $C_{01} = C_{10} = 1$. De façon à nous affranchir du problème

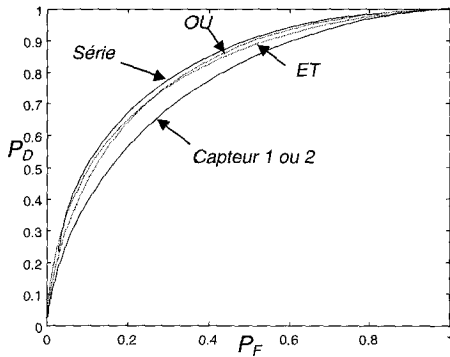


Figure 21. – Courbes COR obtenues en utilisant le critère de Bayes en détection décentralisée série.

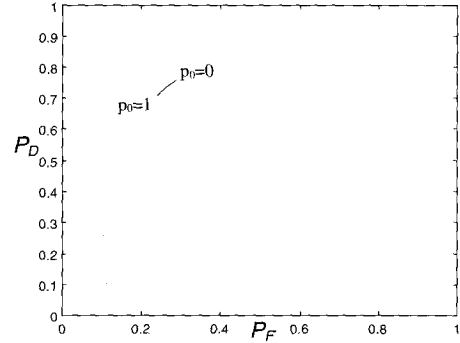


Figure 23. – Courbes COR obtenues en utilisant le critère entropique en détection décentralisée série.

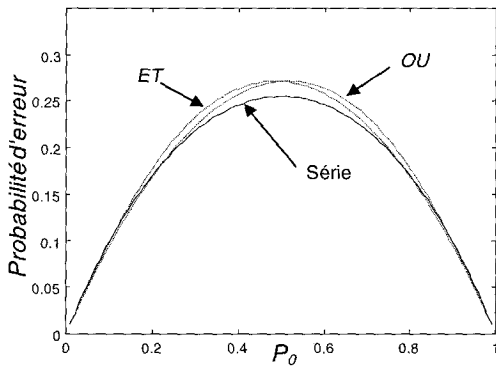


Figure 22. – Courbes de probabilité d'erreur obtenues en utilisant le critère de Bayes en détection décentralisée série.

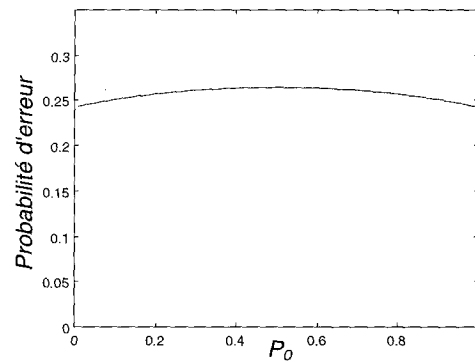


Figure 24. – Courbes de probabilité d'erreur obtenues en utilisant le critère entropique en détection décentralisée série.

de l'ordre des détecteurs, nous avons considéré deux capteurs similaires dont les fonctions de densité de probabilités sont des gaussiennes de moyenne 0 et de variance 1 sous l'hypothèse H_0 , et de variance 1 et de moyenne 1 sous l'hypothèse H_1 .

Les figures 23 et 24 représentent respectivement la courbe COR et la probabilité d'erreur obtenues en minimisant $h(H/u_2)$.

5.4. analyse comparative des méthodes

Pour chaque architecture, les quatre paramètres retenus pour la comparaison des résultats sont P_0 , P_F , P_D et la probabilité d'erreur de prédiction qui est définie par $P_{erreur} = P_0 \cdot P_F + (1 - P_0) \cdot (1 - P_D)$.

Afin d'expliciter l'allure des courbes obtenues, appelons respectivement P_{F_i} et P_{D_i} les valeurs des probabilités P_F et P_D lorsque P_0 tend vers i ($i = 0, 1$). Rappelons que :

- Les valeurs maximales et minimales de P_D et P_F sont obtenues lorsque $P_0 \rightarrow 0$ et $P_0 \rightarrow 1$ en tenant compte du seuil $\lambda = \frac{P_0 \cdot C_{10} - C_{00}}{1 - P_0 \cdot C_{01} - C_{11}}$. A titre d'exemple, le critère de Bayes (considérant des coûts constants $C_{ij} > C_{ii}$ avec $i \neq j$) permet le calcul de ces limites :

Lorsque $P_0 \rightarrow 0$ la règle devient : $\frac{p(y/H_1)}{p(y/H_0)} \underset{u_0=0}{\underset{u_0=1}{\geq}} 0$, ce qui permet de conclure que :

$$P_D = P(u_0 = 1/H_1) \rightarrow P_{D_0} = 1$$

$P(u_0 = 0/H_0) \rightarrow 0$ ce qui entraîne que $P_F = P(u_0 = 1/H_0) \rightarrow P_{F_0} = 1$

De même, lorsque $P_0 \rightarrow 1$ la règle devient :

$$\frac{p(y/H_1)}{p(y/H_0)} \underset{u_0=0}{\underset{u_0=1}{\geq}} +\infty,$$

et donc :

$$P_D = P(u_0 = 1/H_1) \rightarrow P_{D_1} = 0$$

$P(u_0 = 0/H_0) \rightarrow 1$ ce qui entraîne que $P_F = P(u_0 = 1/H_0) \rightarrow P_{F_1} = 0$

- Le seuil λ est une fonction croissante de P_0 . P_D et P_F sont des fonctions décroissantes de P_0 .

A la vue des résultats, quelle que soit l'architecture retenue (centralisée, décentralisée parallèle avec optimisation en deux étapes, décentralisée parallèle avec optimisation simultanée, décentralisée série), les remarques suivantes peuvent être formulées :

- Par définition, la probabilité d'erreur de prédiction est minimisée par le critère de Bayes en utilisant des coûts constants. L'optimisation entropique consistant à utiliser une fonction d'utilité non

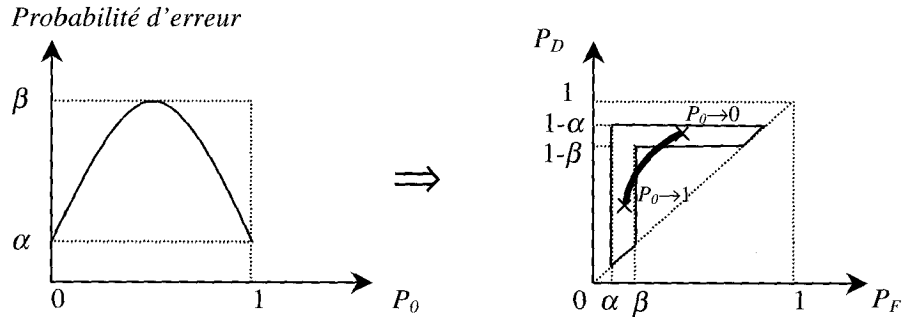


Figure 25. – Passage de la courbe de probabilités d’erreur à la courbe COR.

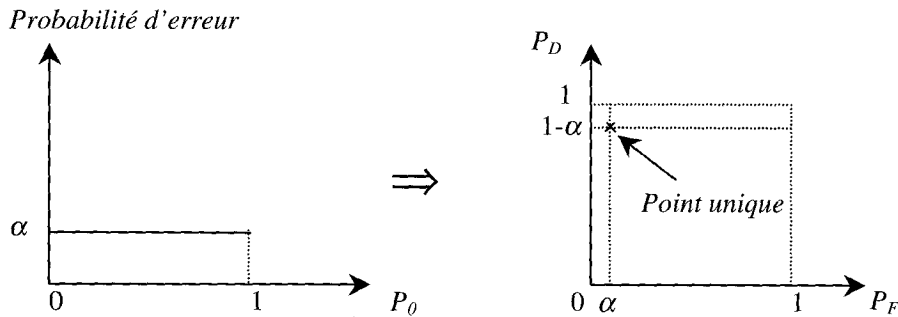


Figure 26. – La courbe COR se limite à un seul point lorsque la probabilité d’erreur est constante.

linéaire, ou linéaire avec des coûts non constants (voir paragraphe 2.2), ne permet pas, en général, d’obtenir de meilleurs résultats sur cet exemple.

• Pourtant, dans le cas particulier de deux hypothèses équiprobables (*i.e.* $P_0 = 0.5$), les deux critères donnent des résultats similaires. La probabilité d’erreur de prédiction est alors maximale ($\lambda = 1$ *via* le critère de Bayes).

• La probabilité d’erreur de prédiction tend vers 0 au fur et à mesure que $P_0 \rightarrow 0$ ou 1 (événements rares) avec le critère de Bayes et les courbes COR s’éloignent notablement du point idéal ($P_F = 0, P_D = 1$). Ce résultat s’explique par le fait que lorsque $P_0 \rightarrow 1$ (*resp.* lorsque $P_0 \rightarrow 0$), le seuil λ prend des valeurs très grandes (*resp.* proches de 0). Dans ce cas de figure ($P_0 \simeq 0$ ou 1), le critère entropique donne une solution peu satisfaisante en termes de probabilités d’erreur de prédiction. En contrepartie, les probabilités de détection et de fausse alarme obtenues restent proches du point idéal ($P_F = 0, P_D = 1$).

L’analyse des courbes de probabilités d’erreur de prédiction obtenues par optimisation entropique semble plus délicate car les coûts ne sont plus constants. En revanche, à partir de celles-ci, nous pouvons expliquer l’allure des courbes COR :

De façon générale, la probabilité d’erreur de prédiction est comprise entre deux bornes α et β (figure 25) :

$$0 \leq \alpha \leq P_{Erreur} = P_0 \cdot P_F + (1 - P_0) \cdot (1 - P_D) \leq \beta \leq 1 \quad (67)$$

(dans le cas particulier de l’utilisation du critère de Bayes, nous avons montré que $\alpha = 0$).

Lorsque $P_0 \neq 1$, cette inégalité peut également s’écrire :

$$\begin{aligned} - \frac{P_0 \cdot (1 - P_F)}{1 - P_0} &\leq \frac{P_0 \cdot P_F - \beta}{1 - P_0} + 1 \leq P_D \\ &\leq \frac{P_0 \cdot P_F - \alpha}{1 - P_0} + 1 \leq \frac{P_0 \cdot P_F}{1 - P_0} + 1 \quad (68) \end{aligned}$$

Les extremums de la courbe COR peuvent alors être encadrés en considérant :

- la limite P_{D_0} lorsque $P_0 \rightarrow 0$ de P_D (inégalité (68)) : $1 - \beta \leq P_{D_0} \leq 1 - \alpha$ avec la contrainte $P_{F_0} \leq P_{D_0}$.
- la limite P_{F_1} lorsque $P_0 \rightarrow 1$ de P_F (inégalité (67)) : $\alpha \leq P_{F_1} \leq \beta$ avec la contrainte $P_{F_1} \leq P_{D_1}$.

Lorsque $\alpha \neq 0$ (cas de l’optimisation entropique), alors $P_D < 1$ et $P_F > 0$ (de façon stricte) avec $P_F \leq P_D$. Cette remarque permet l’explication de la forme des courbes COR obtenues en utilisant le critère entropique (figure 25) constituées d’un arc plus restreint que celui obtenu avec le critère de Bayes.

Lorsque $\alpha = \beta$, c’est-à-dire lorsque la probabilité d’erreur de prédiction reste constante quelle que soit P_0 , alors $P_{F_1} = \alpha$ et $P_{D_0} = 1 - \alpha$. Nous obtenons également la relation $P_D = \frac{P_0 \cdot P_F - \alpha}{1 - P_0} + 1$ lorsque $P_0 \neq 1$. Nous pouvons en déduire que

$$P_{D_1} = \frac{P_0 \cdot P_{F_1} - \alpha}{1 - P_0} + 1 = \frac{P_0 \cdot \alpha - \alpha}{1 - P_0} + 1 = 1 - \alpha = P_{D_0} \cdot P_D$$

reste donc constante, et la relation $P_D = \frac{P_0 \cdot P_F - \alpha}{1 - P_0} + 1$ implique que P_F reste également constante. La courbe COR se limite dans ce cas à un seul point ($P_F = \alpha$, $P_D = 1 - \alpha$) (figure 26).

En particulier, lorsque $\alpha = \beta = 0$ (probabilité d'erreur de prédiction nulle quelle que soit P_0), la courbe COR se limite au point idéal ($P_F = 0$, $P_D = 1$).

Enfin, de façon générale, plus α et β sont proches (la probabilité d'erreur de prédiction tend vers une constante), et plus la courbe COR se rapproche d'un point unique. C'est ce que l'on observe lors de l'optimisation entropique de l'architecture décentralisée série (voir figures 23 et 24).

L'utilisation du critère entropique permet donc de diminuer la plage des valeurs potentielles de P_F et de P_D , et constitue ainsi un critère robuste vis-à-vis des variations de P_0 . En particulier, lorsque P_0 est proche de 0 ou de 1, les points de la courbe COR restent proches du point idéal ($P_F = 0$, $P_D = 1$), ce qui n'est pas le cas lorsqu'on utilise le critère Bayésien. Cela signifie que lorsqu'un événement n'arrive que très rarement (détection de pannes par exemple), l'utilisation du critère de Bayes aura pour effet de le négliger; tandis qu'avec le critère entropique, cet événement est pris en compte de façon non négligeable, au détriment de la probabilité d'erreur de prédiction.

Afin de conclure cette analyse comparative, nous pouvons également remarquer les points suivants :

– **En détection centralisée :** Conformément à notre idée intuitive, les courbes de probabilité d'erreur de prédiction obtenues avec le critère de Bayes ou avec le critère entropique (figures 10 et 12) montrent que le capteur Y_2 est bien meilleur que le capteur Y_1 . Ce résultat se retrouve sur les courbes COR : la courbe COR du capteur Y_2 se trouve plus proche du point idéal que celle du capteur Y_1 .

– **En détection décentralisée parallèle avec optimisation en deux étapes :** Avec le critère de Bayes, la décision prise par le capteur Y_2 (figure 14, fonction f_6) est optimale, quelle que soit la valeur prise par P_0 . Par contre, en utilisant le critère entropique (Figure 16), la combinaison sous la forme d'un « et logique » (resp. d'un « ou logique ») dans le cas où P_0 est proche de 1 (resp. de 0) permet d'obtenir une probabilité d'erreur proche de la probabilité optimale obtenue avec le critère de Bayes.

– **En détection décentralisée parallèle avec optimisation simultanée :** Les figures 18 et 20 permettent de mettre en évidence le fait que la règle de fusion optimale dépend de la probabilité P_0 : pour $0 < P_0 < 0.5$ (resp. pour $0.5 < P_0 < 1$), le « ou logique » (resp. le « et logique ») est optimal. Dans ce cas, la solution entropique est très proche de la solution optimale Bayésienne.

Ces courbes mettent également en évidence le fait que l'optimisation simultanée permet d'obtenir de meilleurs résultats que l'optimisation séparée sur cet exemple.

6. conclusions

Dans le contexte de la théorie de la détection, les critères de Bayes et de Neyman-Pearson sont classiquement employés dans la littérature. Nous avons proposé l'utilisation d'un critère entropique basé sur l'optimisation de la quantité d'information pertinente pour le problème de détection. Nous montrons que le critère entropique consiste à utiliser une fonction d'utilité non linéaire de P_0 , P_F et P_D , ou une fonction d'utilité linéaire avec des coûts non constants dépendant des probabilités *a posteriori*.

Les résultats de l'optimisation entropique des architectures de détection centralisée ont été rappelés. Dans le cadre de la détection décentralisée parallèle, nous avons proposé d'optimiser les détecteurs locaux sans prendre en compte le problème de la fusion, que nous optimisons séparément. L'intérêt principal de cette approche est qu'elle offre une solution simple à mettre en œuvre. Nous avons ensuite rappelé les résultats de l'optimisation simultanée des détecteurs locaux et de l'opérateur de fusion. Nous avons également proposé l'utilisation de ce critère dans le cadre de l'optimisation d'une architecture décentralisée série.

Enfin, une comparaison des résultats de l'optimisation de toutes ces architectures sur un exemple illustratif a permis la mise en évidence de propriétés très intéressantes de robustesse de l'approche informationnelle vis-à-vis d'événements rares, ou vis-à-vis d'événements pour lesquels les probabilités *a priori* sont mal connues. En particulier, l'utilisation du critère entropique n'entraîne pas un éloignement de la courbe COR du point idéal ($P_F = 0$, $P_D = 1$) lorsque P_0 est proche de 0 ou de 1, contrairement au critère de Bayes. Cela signifie que lorsqu'un événement n'arrive que très rarement (détection de pannes par exemple), l'utilisation du critère de Bayes aura pour effet de le négliger; tandis qu'avec le critère entropique, cet événement est pris en compte de façon non négligeable, au détriment de la probabilité d'erreur de prédiction.

Annexe 1

Démontrons que la dérivée de la fonction $f(v) = \alpha \log \frac{\alpha}{\alpha + \beta} +$

$\beta \log \frac{\beta}{\alpha + \beta}$ par rapport à v peut se mettre sous la forme :

$$\frac{\partial f(v)}{\partial v} = \frac{\partial \alpha}{\partial v} \log \frac{\alpha}{\alpha + \beta} + \frac{\partial \beta}{\partial v} \log \frac{\beta}{\alpha + \beta} :$$

$$\begin{aligned} \frac{\partial f(v)}{\partial v} &= \frac{\partial \alpha}{\partial v} \log \frac{\alpha}{\alpha + \beta} + \frac{\alpha}{\alpha} \frac{\partial \alpha}{\partial v} - \frac{\alpha}{\alpha + \beta} \left(\frac{\partial \alpha}{\partial v} + \frac{\partial \beta}{\partial v} \right) \\ &+ \frac{\partial \beta}{\partial v} \log \frac{\beta}{\alpha + \beta} + \frac{\beta}{\beta} \frac{\partial \beta}{\partial v} - \frac{\beta}{\alpha + \beta} \left(\frac{\partial \alpha}{\partial v} + \frac{\partial \beta}{\partial v} \right) \\ \frac{\partial f(v)}{\partial v} &= \frac{\partial \alpha}{\partial v} \left(\log \frac{\alpha}{\alpha + \beta} + 1 - \frac{\alpha}{\alpha + \beta} - \frac{\beta}{\alpha + \beta} \right) \\ &+ \frac{\partial \beta}{\partial v} \left(\log \frac{\beta}{\alpha + \beta} + 1 - \frac{\alpha}{\alpha + \beta} - \frac{\beta}{\alpha + \beta} \right) \end{aligned}$$

Tableau 3. – Les 16 fonctions de fusion potentielles pour 2 capteurs en parallèle.

u_1	u_2	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

$$\frac{\partial f(v)}{\partial v} = \frac{\partial \alpha}{\partial v} \log \frac{\alpha}{\alpha + \beta} + \frac{\partial \beta}{\partial v} \log \frac{\beta}{\alpha + \beta}$$

cqfd.

Annexe 2 : Fonctions de fusion monotones

Dans le cas particulier de la détection décentralisée avec deux capteurs en parallèle, les 16 fonctions de fusion sont données dans le tableau 3.

Quelques-unes de ces 16 fonctions peuvent *a priori* être écartées en considérant la propriété de monotonie. Celle-ci consiste à dire qu’il n’y a aucune raison pour que la décision finale soit $u_0 = 0$ avec un ensemble S de détecteurs locaux décidant $u_i = 1$ alors que la décision finale est $u_0 = 1$ avec un sous-ensemble S' de S de détecteurs locaux décidant $u_i = 1$.

Afin d’être optimal, en supposant que l’on ait $P_{D_i} \geq P_{F_i}$ ($i = 1, \dots, N$), l’opérateur de fusion doit donc nécessairement satisfaire la condition suivante :

$$\forall u^* = (u_1^*, \dots, u_N^*) \text{ et } \forall u^{*'} = (u_1^{*'}, \dots, u_N^{*'}), \text{ si } \forall i \in \{1, \dots, N\} u_i^* \geq u_i^{*' } \text{ alors } f(u^*) \geq f(u^{*' })$$

où :

- $u^* = (u_1^*, \dots, u_N^*)$ est un vecteur de décisions prises par les détecteurs locaux.

- $u^{*' } = (u_1^{*' }, \dots, u_N^{*' })$ est un autre vecteur de décisions prises par les détecteurs locaux.

Cette condition nous permet d’écartier $f_3, f_5, f_7, f_9, f_{10}, f_{11}, f_{12}, f_{13}, f_{14}$ et f_{15} des fonctions de fusions admissibles, pour obtenir le tableau ci-dessous.

Tableau 4. – Les fonctions de fusion admissibles pour 2 capteurs en parallèle.

u_1	u_2	f_1	f_2	f_4	f_6	f_8	f_{16}
0	0	0	0	0	0	0	1
0	1	0	0	0	1	1	1
1	0	0	0	1	0	1	1
1	1	0	1	1	1	1	1

Annexe 3

Démontrons que, de façon générale, pour tout récepteur à rapport de vraisemblance, on a $\frac{\partial P_D}{\partial P_F} = \lambda$, où λ est le seuil permettant de prendre la décision u_0 :

Par définition :

$$\begin{aligned} P_F &= P(u_0 = 1/H_0) = P(\Lambda \geq \lambda/H_0) \\ &= \int_{\lambda}^{+\infty} p(\Lambda/H_0) d\Lambda = A(\lambda) \\ P_D &= P(u_0 = 1/H_1) = P(\Lambda \geq \lambda/H_1) \\ &= \int_{\lambda}^{+\infty} p(\Lambda/H_1) d\Lambda = B(\lambda) \end{aligned}$$

Les probabilités de fausse alarme et de détection s’expriment en fonction du seuil λ fixé. L’élimination de ce seuil entre les deux expressions conduit à une relation liant les probabilités de détection et de fausse alarme : $\lambda = A^{-1}(P_F)$ et $P_D = B(A^{-1}(P_F))$, par conséquent $P_D = f(P_F)$. Les courbes COR utilisées au paragraphe 5 pourront donc être tracées en utilisant λ comme paramètre.

• Lemme : On a toujours $\frac{P(\Lambda = \lambda/H_1)}{P(\Lambda = \lambda/H_0)} = \lambda$

De plus :

$$P_F(\lambda) = \int_{\lambda}^{+\infty} p(\Lambda/H_0) d\Lambda \Rightarrow \left(\frac{\partial P_F}{\partial \Lambda} \right)_{\Lambda=\lambda} = -P(\Lambda = \lambda/H_0)$$

car $\lim_{\Lambda \rightarrow +\infty} p(\Lambda/H_0) = 0$

et :

$$P_D(\lambda) = \int_{\lambda}^{+\infty} p(\Lambda/H_1) d\Lambda \Rightarrow \left(\frac{\partial P_D}{\partial \Lambda} \right)_{\Lambda=\lambda} = -P(\Lambda = \lambda/H_1)$$

Donc, d’après le lemme : $\frac{\left(\frac{\partial P_D}{\partial \Lambda} \right)_{\Lambda=\lambda}}{\left(\frac{\partial P_F}{\partial \Lambda} \right)_{\Lambda=\lambda}} = \frac{P(\Lambda = \lambda/H_1)}{P(\Lambda = \lambda/H_0)} = \lambda$

Et on en déduit que : $\frac{\partial P_D}{\partial P_F} = \lambda$

cqfd.

BIBLIOGRAPHIE

[1] P.K. Varshney, *Distributed Detection and Data Fusion*, Springer Verlag, New-York, 1997.
 [2] D. Middleton, *Statistical Communication Theory*, McGraw-Hill, New-York, 1960.

- [3] T.L. Gabrielle, « *Information Criteria for Threshold Determination* », *IEEE Trans. on Information Theory*, vol.6, October 1966, pp.484-486.
- [4] I.Y. Hoballah and P.K. Varshney, « *An Information Theoretic Approach to the Distributed Detection Problem* », *IEEE Trans. on Information Theory*, vol.IT-35, n°5, September 1989, pp.988-994.
- [5] D. Warren, P. Willet, and R. Rampertab, « *Shannon's Information in Decentralized Signal Detection* », *Proc. 23rd Conference on Info. Sciences and Systems*, Baltimore, March 1989.
- [6] H. Ventsel, *Théorie des probabilités*, Ed. Mir Moscou, 1973.
- [7] R. Ash, *Information Theory*, John Wiley and Sons, 1965.
- [8] R.R. Tenney and N.R. Sandell, « *Detection with Distributed Sensors* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-17, n°4, July 1981, pp.98-101.
- [9] F.A. Sadjadi, « *Hypotheses Testing in a Distributed Environment* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-22, n°2, March 1986, pp.134-137.
- [10] Z. Chair and P.K. Varshney, « *Optimal Data Fusion in Multiple Sensor Detection Systems* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-22, n°1, January 1986, pp.98-101.
- [11] S.C.A. Thomopoulos, R. Viswanathan, and D.K. Bougoulías, « *Optimal Decision Fusion in Multiple Sensor Systems* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-23, n°5, September 1987, pp.644-653.
- [12] S.C.A. Thomopoulos, R. Viswanathan, and D.K. Bougoulías, « *Optimal Distributed Decision Fusion* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-25, n°5, September 1989, pp.761-765.
- [13] A.R. Reibman and L.W. Nolte, « *Optimal Detection and Performance of Distributed Sensor System* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-23, n°1, January 1987, pp.24-30.
- [14] I.Y. Hoballah, *On the Design and Optimisation of Distributed Signal Detection and Parameter Estimation Systems*, Ph.D. Dissertation, Syracuse University, November 1986.
- [15] I.Y. Hoballah and P.K. Varshney, « *Distributed Bayesian Signal Detection* », *IEEE Trans. on Information Theory*, vol.IT-35, n°5, September 1989, pp.995-1000.
- [16] Z.B. Tang, *Optimisation of Detection Networks*, Ph.D. Dissertation, University of Connecticut, 1990.
- [17] R. Srinivasan, « *Distributed Radar Detection Theory* », *IEEE Proc.*, vol.133, Part F, Inst. of Electrical Engineers, England, UK, February 1986, pp.55-60.
- [18] E. Drakopoulos and C.C. Lee, « *Optimum Multisensor Fusion of Correlated Local Decisions* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-27, n°4, July 1991, pp.593-605.
- [19] P.F. Swaszek, « *On the Performance of Serial Networks in Distributed Detection* », *IEEE Trans. on Aerospace and Electronic Systems*, vol.AES-29, n°1, January 1993, pp.254-260.
- [20] Z.B. Tang, K.R. Pattipati, and D.L. Kleinman, « *Optimisation of Detection Networks : Part I - Tandem Structures* », *IEEE Trans. on Systems, Man, and Cybernetics*, vol.SMC-21, n°5, Sept.-Oct. 1991, pp.1044-1059.

Manuscrit reçu le 11 mars 1999.

LES AUTEURS

Christophe DESROUSSEAUX



Christophe Desrousseau est né en 1971. Ingénieur ISEN, il a soutenu sa thèse de Doctorat en décembre 1998. Son activité de recherche s'est développée autour de la fusion de données et plus particulièrement sur les problèmes d'optimisation des architectures de détection.

Denis POMORSKI



Docteur de l'Université Lille I en Automatique et Informatique Industrielle, Denis Pomorski est actuellement Maître de Conférences à l'IUT A de Lille. Son activité de recherche porte essentiellement sur l'utilisation de la Théorie de l'Information dans le cadre de l'Analyse Structurale des Systèmes Complexes, en vue d'un Apprentissage Automatique Symbolique / Numérique.