

Combinaison d'évidence pour la discrimination en présence d'étiquetage imprécis

Evidence combination for classification with imprecise labels

E. Lefevre²

P. Vannoorenberghe¹

O. Colot²

T. Dencœur¹

¹ HEUDIASYC, UMR 6599 CNRS

Université de Technologie de Compiègne

BP 20529, 60205 Compiègne cedex

² Perception Systèmes Information, EA 2120

Université / INSA de Rouen

Place Emile Blondel, 76821 Mont Saint-Aignan cedex

Patrick.Vannoorenberghe@hds.utc.fr

Résumé

Dans ce papier, on se propose d'appliquer la théorie des fonctions de croyance au problème d'étiquetage imprécis. Ce type de problème d'apprentissage, dans lequel la classe des vecteurs d'apprentissage n'est que partiellement connue, peut être rencontré dans de multiples situations. C'est le cas, notamment quand l'étiquetage est réalisé par plusieurs experts dont les avis divergent. Une technique utilisant les fonctions de croyance a été introduite par T. Dencœur dans [1]. Elle permet de fusionner, par la règle de combinaison de Dempster, des informations issues du voisinage du vecteur à classer dans l'ensemble d'apprentissage. Si ces informations sont en conflit (mauvaise modélisation des fonctions de croyance, nombre élevé de sources d'informations, ...), l'utilisation de la règle de combinaison de Dempster peut être dangereuse. Dans cet article, on se propose de comparer cette règle avec plusieurs règles de combinaison d'évidence qui permettent la prise en compte du conflit. Ces règles de fusion ont été rassemblées récemment [2] sous un même formalisme et, dans une version optimisée, peuvent permettre une redistribution pondérée adéquate du conflit. Des tests sur des données simulées permettent de mettre en évidence les bonnes performances de tels opérateurs vis-à-vis de la règle de combinaison de Dempster.

Mots Clef

Classification, Fonctions de croyance, Apprentissage, Etiquetage imprécis.

Abstract

In this paper, we propose to apply the belief function theory to imprecise labelling. This kind of learning problem, in which the class of learning vectors is partially known, can be encountered in many different situations. That is the case when the training set is labelled by several experts who have conflicting opinions. Using the belief function theory, a method has been introduced by T. Dencœur in [1]. The technique consists in aggregating, with the Dempster rule of combination, information extracted from the neighborhood of the vector to classify in the training set. If these information sources are in conflict, using the Dempster's rule of combination can be dangerous. In this paper, we propose to compare this rule with several evidence based operators which allow to take into account the conflict between sources. These combination rules have been summarized under a general framework and allow to redistribute optimally the conflict. Tests on several simulated datasets demonstrate the good performance of these operators vs. the Dempster rule of combination.

Keywords

Classification, Belief function, Learning, Imprecise labels.

1 Introduction

Dans cet article, on s'intéresse au problème de discrimination qui s'attache à assigner un vecteur d'entrée inconnu \mathbf{x} à une classe w_q qui prend ses valeurs dans $\Omega = \{\omega_1, \dots, \omega_Q\}$. Généralement, le problème

consiste à construire une fonction de décision par la seule connaissance d'un ensemble d'apprentissage \mathcal{L} composé de n vecteurs. Chaque vecteur dans \mathcal{L} est représenté par un vecteur d'attribut \mathbf{x}^i de dimension p et son label correspondant ω^i . Pour construire cette fonction de décision, plusieurs méthodes utilisant la théorie de l'évidence ont été récemment introduites [3, 4]. Elles permettent d'obtenir une information plus riche que les méthodes classiques car elles donnent de manière directe l'incertitude sur la prédiction.

Les exemples dans \mathcal{L} sont généralement étiquetés par un expert du domaine ou de l'application envisagée, mais il arrive parfois que cet étiquetage revête un caractère imprécis. C'est le cas notamment lorsque l'étiquetage est réalisé par plusieurs experts qui ne sont pas en accord. Ainsi, on peut reformuler le problème général de la discrimination en présence d'étiquetage imprécis par ces termes. On suppose que l'ensemble d'apprentissage \mathcal{L} est composé de n paires $(x^i, m_{\mathcal{L}}^i)$ où x^i est un vecteur p -dimensionnel (généralement appelé vecteur-forme) et où $m_{\mathcal{L}}^i$ est une fonction de $2^{\Omega} \rightarrow [0, 1]$ qui vérifie $\sum_{A \subseteq \Omega} m_{\mathcal{L}}^i(A) = 1$ et $m_{\mathcal{L}}^i(\emptyset) = 0$. Par ce biais, l'étiquetage prend la forme d'une fonction de croyance [5]. Dans cet article, on se propose d'étendre la méthode de discrimination présentée dans [3] en comparant la combinaison des sources d'information par la règle de Dempster avec d'autres opérateurs de fusion.

L'article est organisé de la manière suivante. On rappelle tout d'abord dans la section 2 quelques fondamentaux de la théorie de l'évidence qui permettent notamment d'introduire, dans la section 3, les solutions envisagées au problème de discrimination en présence d'étiquetage imprécis. La section 4 permet de présenter la technique d'optimisation des paramètres des méthodes employées. Enfin, quelques résultats sur des données simulées sont présentés dans la section 5.

2 Les fonctions de croyance

Nous rappelons dans cette section quelques éléments théoriques relatifs aux fonctions de croyance. La théorie de l'évidence fut initialement introduite par Dempster [6] lors de ses travaux sur les bornes inférieure et supérieure d'une famille de distributions de probabilités. A partir de ce formalisme mathématique, Shafer [5] a montré l'intérêt des fonctions de croyance pour la modélisation de connaissances incertaines. L'utilité des fonctions de croyance, comme alternative aux probabilités subjectives, a été démontrée plus tard de manière axiomatique par Smets [7, 8] au travers du *Modèle de Croyances Transférables*.

Dans le cadre de la théorie de l'évidence appliquée à

la reconnaissance de formes, l'ensemble Ω est appelé *cadre de discernement*. On suppose que les hypothèses dans Ω sont exclusives et que le cadre de discernement est exhaustif. On définit une masse de probabilité élémentaire, appelée *masse de croyance*, qui caractérise la véracité d'une proposition A pour une source d'information S . La masse m associée à cette source est alors définie par :

$$m : 2^{\Omega} \rightarrow [0, 1] \quad (1)$$

et vérifie les propriétés suivantes : $m(\emptyset) = 0$ et $\sum_{A \subseteq \Omega} m(A) = 1$. Cette fonction se différencie d'une probabilité par le fait que la totalité de la masse de croyance est répartie non seulement sur les hypothèses singletons ω_q , mais aussi sur les hypothèses combinées. La modélisation issue de la fonction m est appelée jeu de masses. A partir de la fonction m , on définit respectivement les fonctions de *crédibilité* Cr et de *plausibilité* Pl par :

$$Cr(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

$$Pl(A) = \sum_{(A \cap B) \neq \emptyset} m(B) = 1 - Cr(\bar{A}) \quad (3)$$

où \bar{A} représente l'événement contraire de la proposition A . La crédibilité $Cr(A)$ mesure la force avec laquelle on croit en la véracité de la proposition. La plausibilité $Pl(A)$, fonction duale de la crédibilité, mesure l'intensité avec laquelle on ne doute pas de A .

Dans le cadre de la théorie de l'évidence de Dempster-Shafer, la fusion des informations issues de sources distinctes est réalisée en utilisant la *loi de combinaison de Dempster*. Celle-ci, qui s'avère commutative et associative, est définie par :

$$\forall A \subseteq \Omega \quad m(A) = m_1(A) \oplus \dots \oplus m_J(A) \quad (4)$$

où \oplus représente l'opérateur de combinaison. Dans un cas à deux sources notées S_i et S_j , la combinaison peut se mettre sous la forme :

$$m(A) = \frac{1}{1 - \kappa} \sum_{(B \cap C) = A} m_i(B).m_j(C) \quad (5)$$

où κ est défini par :

$$\kappa = \sum_{(B \cap C) = \emptyset} m_i(B).m_j(C). \quad (6)$$

Dans l'équation (5), le coefficient κ reflète le conflit existant entre les deux sources S_i et S_j et le quotient $1/(1 - \kappa)$ est un terme de normalisation. Lorsque κ est

égal à 1, les sources sont en conflit total et les informations ne peuvent être fusionnées. Au contraire, lorsque κ est nul, les sources sont en parfait accord. Cette règle de fusion, déduite de la règle de conditionnement [9], a été critiquée dans plusieurs travaux dont [10], en particulier dans le cas de sources en conflit total. Nous reviendrons en détail sur ces notions dans la section suivante.

Pour la prise de décision, on suppose que nous avons une fonction de croyance m sur Ω qui résume l'ensemble des informations apportées. La décision consiste à choisir une action a parmi un ensemble fini d'actions \mathcal{A} . Une fonction de perte $\lambda : \mathcal{A} \times \Omega \rightarrow \mathbb{R}$ est supposée définie de telle manière que $\lambda(a, \omega)$ représente la perte encourue si l'on choisit l'action a lorsque la vraie classe est ω . A partir d'arguments de rationalité, Smets [8] propose de transformer m en une fonction de probabilité P_{Bet} sur Ω (appelée fonction de probabilité pignistique) définie pour tout $\omega \in \Omega$ comme :

$$P_{Bet}(\omega) = \sum_{A \ni \omega} \frac{m(A)}{|A|}, \quad (7)$$

où $|A|$ est la cardinalité de $A \subseteq \Omega$. Dans cette transformation, la masse de croyance $m(A)$ est uniformément distribuée parmi les éléments de A . A partir de cette probabilité, on peut associer à chaque action $a \in \mathcal{A}$ un risque défini comme le coût espéré relatif à P_{Bet} si on choisit a :

$$R(a) = \sum_{\omega \in \Omega} \lambda(a, \omega) P_{Bet}(\omega). \quad (8)$$

L'idée consiste ensuite à choisir l'action qui minimise ce risque, généralement appelé risque pignistique. De la même manière, la prise de décision peut être basée sur une extension non probabiliste du concept mathématique de perte [11]. Par exemple, le concept d'espérance inférieure conduit à la définition du coût espéré minimum comme :

$$R_*(a) = \sum_{A \subseteq \Omega} m(A) \min_{\omega \in A} \lambda(a, \omega), \quad (9)$$

qui permet d'obtenir une stratégie de décision différente. En reconnaissance des formes, $\Omega = \{\omega_1, \dots, \omega_Q\}$ est l'ensemble des classes et les éléments de \mathcal{A} sont généralement les actions a_q qui consistent à assigner le vecteur inconnu à la classe ω_q . Avec des coûts 0-1 ($\lambda(a_q, \omega_r) = 1 - \delta_{q,r}$ pour $q, r \in \{1, \dots, Q\}$), il est démontré [11] que la minimisation du risque pignistique R conduit à choisir la classe ω_0 de plus grande probabilité pignistique. Si une action supplémentaire de rejet a_0 de coût constant λ_0 est possible, alors le vecteur est rejeté si $P_{Bet}(\omega_0) < 1 - \lambda_0$.

3 Solutions envisagées

Une solution au problème d'apprentissage considéré dans ce papier a été introduite par T. Denœux dans [3] dans le cas d'un étiquetage précis mais a été généralisé aux étiquettes imprécises dans [12]. Il est à noter qu'une comparaison avec une approche du type maximum de vraisemblance a été également présentée dans [13].

3.1 Méthode proposée

On suppose que les k -ppv de \mathbf{x} dans l'ensemble d'apprentissage \mathcal{L} sont autant de sources d'informations indépendantes précisant le degré de croyance de \mathbf{x} . Chacune des k fonctions de croyance m^i est construite par affaiblissement en fonction de la distance entre \mathbf{x} et \mathbf{x}^i . Elle est définie $\forall A \in 2^\Omega$ par :

$$m^i(A) = \alpha \phi(d^i) \quad (10)$$

$$m^i(\Omega) = 1 - \alpha \phi(d^i). \quad (11)$$

Dans ces équations, d^i est la distance euclidienne entre \mathbf{x} et \mathbf{x}^i , α est un paramètre et $\phi(\cdot)$ est une fonction décroissante monotone qui vérifie $\phi(0) = 1$ et $\lim_{d \rightarrow \infty} \phi(d) = 0$. Elle peut s'exprimer par :

$$\phi(d^i) = \exp(-\gamma(d^i)^2) \quad (12)$$

avec γ un paramètre positif. Les k fonctions de croyance m^i sont ensuite agrégées en utilisant la règle de combinaison de Dempster :

$$\hat{m} = \bigoplus_{i=1}^k m^i. \quad (13)$$

Si les sources d'information sont en conflit (mauvaise modélisation des fonctions de croyance, nombre élevé de sources d'informations, \dots), il se peut que l'utilisation de la règle de combinaison de Dempster soit dangereuse (voire même impossible). Une solution à ce problème a été introduite dans [14, 2] et est discutée dans le paragraphe suivant.

3.2 Variantes de la méthode

Dans cet article, on se propose de comparer la règle de Dempster avec plusieurs règles de combinaison d'évidence qui permettent la prise en compte du conflit. Ces règles de fusion ont été rassemblées récemment [2] sous un même formalisme et, dans une version optimisée, peuvent permettre une redistribution pondérée adéquate du conflit. Le conflit est redistribué de manière pondérée entre les sous-ensembles de Ω . L'opérateur peut ainsi se mettre sous la forme :

$$\forall A \subseteq \Omega \quad m(A) = m_{\cap}(A) + w(A, m^i) \cdot \kappa. \quad (14)$$

Dans cette équation, m_{\cap} est la combinaison conjonctive (la règle de Dempster non normalisée) des k fonctions de croyance $m_{\cap} = \bigcap m^i$, κ est le conflit total entre les fonctions de croyance et w est un poids qui dépend de A et des fonctions de croyance m^i . Le conflit peut être généré par un défaut de fiabilité d'une partie des sources d'information. Cet argument a été repris dans le cadre des opérateurs présentés par Yager [10] et par Dubois et Prade [15].

Dans le cas de l'opérateur de Yager [10], on suppose que l'une des sources intervenant dans la combinaison est fiable. Ainsi, la solution est obligatoirement dans le référentiel. Mais ne sachant quelle source donne la vraie solution, Yager propose d'attribuer la masse conflictuelle $m(\emptyset)$ à l'ensemble Ω . La masse résultante, m_Y , de cette combinaison, pour deux sources d'information $\{S_i, S_j\}$, est obtenue de la manière suivante :

$$\begin{cases} m_Y(A) &= m_{\cap}(A) & \forall A \subset \Omega \\ m_Y(\Omega) &= m_{\cap}(\Omega) + \kappa. \end{cases} \quad (15)$$

La combinaison proposée par Dubois et Prade [15], dans le cadre de la fusion de deux sources d'information $\{S_i, S_j\}$, peut s'expliquer de la manière suivante. Soit S_i une source soutenant la proposition B avec une masse de croyance $m_i(B)$ et soit une source S_j soutenant la proposition C avec une masse de croyance $m_j(C)$. Lorsque les propositions soutenues par ces deux sources sont contradictoires, et ne sachant pas quelle source est fiable, le principe de minimum de spécificité impose de redistribuer la masse associée à cette contradiction $m_i(B).m_j(C)$, sur l'union des propositions c'est-à-dire $(B \cup C)$. L'opérateur de Dubois et Prade est alors défini de la manière suivante pour deux sources d'information :

$$m_D(A) = m_{\cap}(A) + \sum_{\substack{B \cup C = A \\ B \cap C = \emptyset}} m_i(B).m_j(C) \quad \forall A \neq \emptyset \quad (16)$$

et :

$$m_D(\emptyset) = 0 \quad (17)$$

Dans [2], il est montré que l'opérateur proposé dans l'équation (14) est suffisamment générique pour permettre de retrouver les opérateurs précédemment cités sous réserve de choisir les poids $w(A, m^i)$ de manière adéquate.

4 Optimisation des paramètres

L'optimisation des paramètres de la méthode précédemment introduite a été réalisé par T. Dencoux dans [3] en minimisant le critère d'erreur suivant :

$$E = \sum_{i=1}^n \sum_{q=1}^Q (\hat{P}_{Bet}^i(\{\omega_q\}) - u_q^i)^2 \quad (18)$$

TAB. 1 – Propriétés pour le critère

1	$C(\omega_1, \omega_2)$	$C = \begin{cases} 1 & \text{si } \omega_1 = \omega_2 \\ 0 & \text{sinon} \end{cases}$
2	$C(A, B)$	$C = 1 - Incl(A, B)$
3	$C(\omega, A)$	$C = \begin{cases} 0 & \text{si } \omega \in A \\ 1 & \text{sinon} \end{cases}$
4	$C(\hat{m}, \Omega)$	$C = 0$

où u_q^i est l'indicateur d'appartenance du vecteur \mathbf{x}^i ($u_q^i = 1$ si $\omega^i = \omega_q$), et $\hat{P}_{Bet}^i(\{\omega_q\})$ est la probabilité pignistique de ω_q pour le vecteur \mathbf{x}^i donnée par la méthode des distances décrite au paragraphe 3.1. Ce critère est défini dans le cas d'étiquettes précises. Dans le cas où l'étiquetage est imprécis, il est tout à fait envisageable d'adapter le critère présenté à l'équation (18) aux étiquettes données sous formes de fonctions de croyance. Ainsi, dans cet article, on propose d'utiliser le critère introduit dans [12] qui est défini par :

$$C(\hat{m}, m_{\mathcal{L}}) = n - \sum_{i=1}^n \sum_{A \subseteq \Omega} \hat{P}_{Bet}^i(A).m_{\mathcal{L}}^i(A). \quad (19)$$

Il est à noter que la minimisation de ce critère peut être réalisée par le biais des poids $w(A, m^i)$ qui interviennent dans la fonction de croyance \hat{m} délivrée par la règle de décision. Même si la définition du critère ne constitue pas l'objectif de nos travaux, il est intéressant de remarquer que le critère précédemment cité répond à quelques propriétés. Ainsi, nous avons récapitulé 4 propriétés dans le tableau 1 qui sont à satisfaire dans le cas d'étiquettes imprécises. La propriété 1 est l'extension des coûts 0 – 1 aux étiquettes précises. La propriété 2 concerne l'adéquation du coût pour les étiquettes imprécises. Enfin, la propriété 4 concerne un vecteur qui ne serait pas étiqueté dans l'ensemble d'apprentissage. On peut remarquer que l'ajout de vecteurs dont l'étiquetage correspond à $m_{\mathcal{L}}(\Omega) = 1$ ne modifie pas la valeur de C . Il est envisageable d'utiliser d'autres critères qui pourraient répondre à ces conditions [16].

5 Simulations

Pour illustrer le comportement de la méthode en fonction de l'opérateur de combinaison choisi, on génère un ensemble d'apprentissage \mathcal{L} composé de 3 classes gaussiennes de 100 vecteurs chacune, issues des distributions suivantes $\mathcal{N}(\mu_q, \Sigma_q)$:

$$\mu_1 = (0, 0)^t, \quad \mu_2 = (2, 2)^t, \quad \mu_3 = (10, 10)^t \quad (20)$$

$$\Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \Sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (21)$$

A partir de ces distributions, on réalise un étiquetage imprécis des vecteurs de l'ensemble d'apprentissage de deux façons différentes permettant ainsi d'illustrer deux situations différentes. L'un correspond à un étiquetage ensembliste au sens strict du terme tandis que l'autre est un étiquetage imprécis qui prend en compte la fiabilité de l'expert.

5.1 Etiquetage ensembliste

Pour le vecteur \mathbf{x}^i de classe précise ω^i dans l'ensemble d'apprentissage, l'étiquetage imprécis est généré de la manière suivante. On tire aléatoirement un sous-ensemble A tel que $\omega^i \subset A \subseteq \Omega$. La totalité de la masse de croyance est affectée à ce sous-ensemble :

$$m_{\mathcal{L}}^i(A) = 1. \quad (22)$$

L'ensemble de test \mathcal{T} est composé de 250 vecteurs par classe. Les paramètres de la méthode sont fixés de la manière suivante : $k = 15$, $\alpha = 0.99$ et le paramètre γ et pris comme préconisé dans [1] égale à la moyenne des distances aux k voisins dans l'ensemble d'apprentissage.

Les figures 1, 3 et 5 représentent, respectivement, les probabilités pignistiques pour les trois règles de combinaison envisagées (la règle de combinaison de Dempster, l'opérateur de Yager et l'opérateur pondéré avec optimisation des poids). Dans cet exemple à 3 classes, il est possible de représenter les probabilités pignistiques dans un triangle équilatéral dont chacune des hauteurs représente la probabilité pignistique de chacune des classes. De cette manière, les vecteurs de l'ensemble de test sont représentés dans le triangle ($\square = \omega_1$, $\circ = \omega_2$, $+$ = ω_3) ainsi que la zone de rejet possible pour deux valeurs de coût ($\lambda_0 = 0.25$ et $\lambda_0 = 0.5$) en trait plein. La prise de décision sans rejet est également représentée en trait pointillé dans ce même triangle.

En terme de classification, on a représenté sur les figures 2 et 4, le taux d'erreur en classification en fonction du taux de rejet pour les deux règles de décision qui correspondent à la minimisation du risque pignistique (cf. équation (8)) et du risque inférieur (cf. équation (9)). On peut remarquer que la règle de combinaison de Dempster obtient de moins bonnes performances que les deux autres opérateurs. Pour l'opérateur pondéré, la répartition des poids est la suivante : $w(\{\omega_1, \omega_2\}) = 0.41$, $w(\{\omega_1, \omega_3\}) = 0.1$, $w(\{\omega_2, \omega_3\}) = 0.13$ et $w(\Omega) = 0.36$.

5.2 Etiquetage imprécis

Dans cette seconde simulation, on souhaite un étiquetage imprécis (c'est-à-dire crédibiliste) pour chacun des vecteurs \mathbf{x}^i . Pour cela, on tire aléatoirement une probabilité δ qui suit une distribution uniforme sur $[0, 1]$. Avec une probabilité δ , on tire un sous-ensemble $A \subseteq \Omega$ qui contienne la vraie classe ω^i ($\omega^i \subseteq A \subseteq \Omega$). Dans ce cas, le sous-ensemble A ne contient pas forcément (avec une probabilité $1 - \delta$) la vraie classe ω^i . Enfin, on répartit la masse de croyance de la manière suivante :

$$m_{\mathcal{L}}^i(A) = \delta \quad (23)$$

$$m_{\mathcal{L}}^i(\Omega) = 1 - \delta. \quad (24)$$

De cette façon, la fonction de croyance $m_{\mathcal{L}}^i$ représente la confiance dans la vraie classe ω^i à laquelle on ajoute le degré de fiabilité δ de l'expert qui réalise l'étiquetage. De la même façon que précédemment, on a représenté les probabilités pignistiques pour les trois opérateurs (figures 6, 8 et 10), ainsi que les performances en classification pour les deux règles de décision (figures 7 et 9). Pour cette simulation, les poids $w(A, m^i)$ ont été fixés en répartissant l'unité de manière uniforme ($w(A, \cdot) = 1/(2^{|\Omega|} - 1) \quad \forall A \subseteq \Omega$). On constate donc que l'opérateur de Yager et l'opérateur proposé ont des comportements similaires au niveau des probabilités pignistiques. Ce constat n'est plus vrai pour les plausibilités et dans cette simulation, c'est l'opérateur pondéré qui obtient les meilleures performances.

6 Conclusion

Dans cet article, on s'est intéressé au problème de discrimination en présence d'étiquettes imprécises, par une technique qui tire profit du formalisme de la théorie des fonctions de croyance. La méthode combine, par la règle de combinaison de Dempster, des fonctions de croyance issues de la modélisation à partir des données de l'ensemble d'apprentissage. Nous avons comparé cette règle de fusion en introduisant plusieurs règles de combinaison d'évidence qui peuvent se formaliser par une fusion pondérée des différentes sources en présence. Une optimisation de ces pondérations a été introduite en minimisant un critère d'erreur. Les simulations réalisées dans le cadre de cet article mettent en évidence qu'un choix judicieux de la règle de combinaison peut permettre d'obtenir de meilleures performances en terme de classification. Ainsi, l'utilisation de l'opérateur proposé dans [2, 17] est une solution élégante qui permet non seulement de retrouver l'ensemble des opérateurs introduits dans la littérature mais aussi de créer une famille beaucoup plus générique pour la gestion du conflit.

Références

- [1] T. Denoeux, "A k-nearest neighbour classification rule based on Dempster-Shafer theory," *IEEE Transactions on Systems Man and Cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.
- [2] E. Lefevre, O. Colot, P. Vannoorenberghe, and D. Bruccq, "Informations et combinaison : Les liaisons conflictuelles," *Traitement du Signal*, vol. A Paraitre, pp. xx–xx, 2001.
- [3] L. Zouhal and T. Denoeux, "An evidence-theoretic k-nn rule with parameter optimization," *IEEE Transactions on Systems, Man and Cybernetics-Part C*, vol. 28, pp. 263–271, May 1998.
- [4] A. Appriou, *Aggregation and Fusion of Imperfect Information*, ch. Uncertain Data Aggregation in Classification and Tracking Processes, pp. 231–260. Physica-Verlag, 1998.
- [5] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [6] A. Dempster, "Upper and lower probabilities induced by multivalued mapping," *Annals of Mathematical Statistics*, vol. AMS-38, pp. 325–339, 1967.
- [7] P. Smets, "What is Dempster-Shafer's model?," in *Advances in the Dempster-Shafer Theory of Evidence* (R. Yager, M. Fedrizzi, and J. Kacprzyk, eds.), pp. 5–34, Wiley, 1994.
- [8] P. Smets and R. Kennes, "The Transferable Belief Model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [9] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 447–458, 1990.
- [10] R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information Sciences*, vol. 41, pp. 93–138, 1987.
- [11] T. Denoeux, "Analysis of evidence-theoretic decision rules for pattern classification," *Pattern Recognition*, vol. 30, no. 7, pp. 1095–1107, 1997.
- [12] T. Denoeux and L. Zouhal, "Handling possibilistic labels in pattern classification using evidential reasoning," *Fuzzy Sets and Systems*, vol. 122, no. 3, pp. 47–62, 2001.
- [13] C. Ambroise, T. Denoeux, G. Govaert, and P. Smets, "Learning from an imprecise teacher : Probabilistic and evidential approaches," in *Applied Stochastic Models and Data Analysis*, pp. 100–105, 2001.
- [14] E. Lefevre, O. Colot, P. Vannoorenberghe, and D. D. Bruccq, "A generic framework for resolving the conflict in the combination of belief function," in *Proceedings of FUSION'2000, Third International Conference on Information Fusion*, (Paris, France), pp. MOD4 11–18, July 2000.
- [15] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Comput. Intell.*, vol. 4, pp. 244–264, 1988.
- [16] A. Jousselme, D. Grenier, and E. Bossé, "A new distance between two bodies of evidence," *Information fusion*, vol. 2, pp. 91–101, 2001.
- [17] E. Lefevre, P. Vannoorenberghe, and O. Colot, "Combination rules for belief functions," in *Applied Stochastic Models and Data Analysis*, pp. 678–683, 2001.

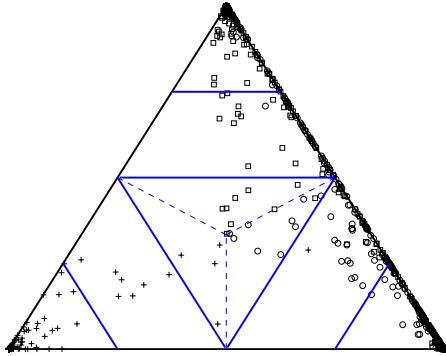


FIG. 1 – Probabilité pignistique avec l'opérateur de Dempster, $\square = \omega_1$, $\circ = \omega_2$, $+$ $= \omega_3$

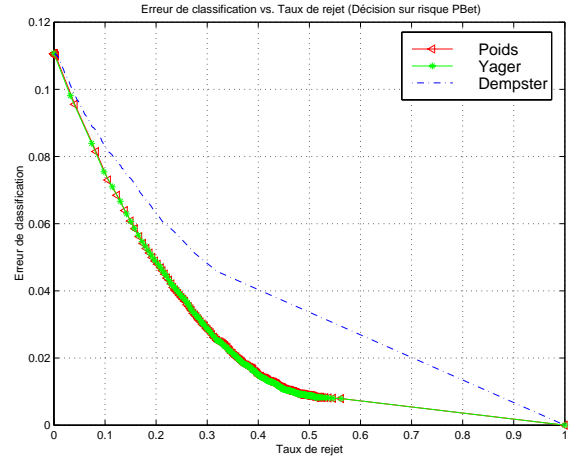


FIG. 2 – Taux d'erreur vs. taux de rejet pour le risque pignistique, $\cdot -$ = Dempster, $*$ = Yager, \triangleleft = Poids

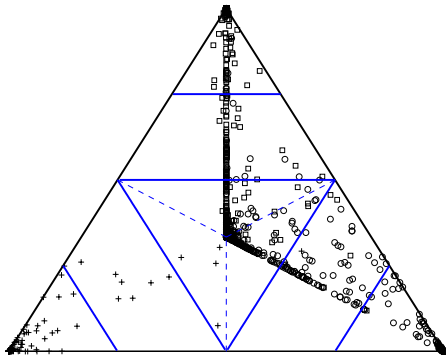


FIG. 3 – Probabilité pignistique avec l'opérateur de Yager, $\square = \omega_1$, $\circ = \omega_2$, $+$ $= \omega_3$

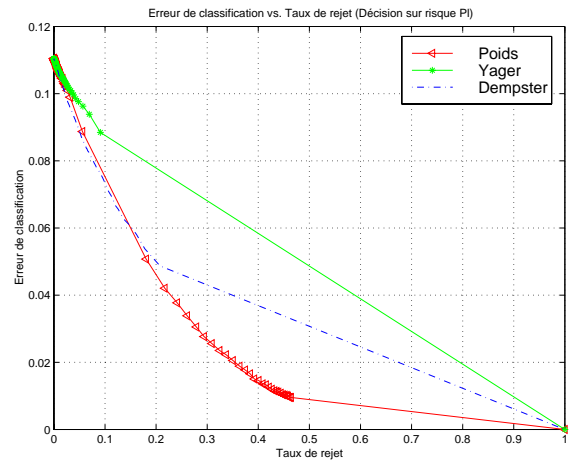


FIG. 4 – Taux d'erreur vs. taux de rejet pour le risque inférieur, $\cdot -$ = Dempster, $*$ = Yager, \triangleleft = Poids

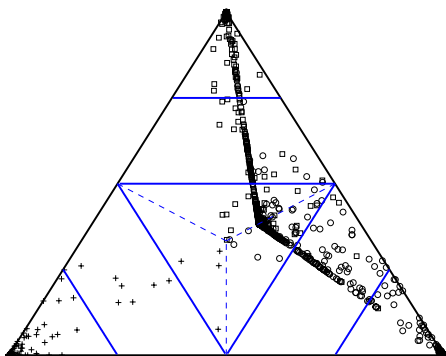


FIG. 5 – Probabilité pignistique avec l'opérateur pondéré, $\square = \omega_1$, $\circ = \omega_2$, $+$ $= \omega_3$

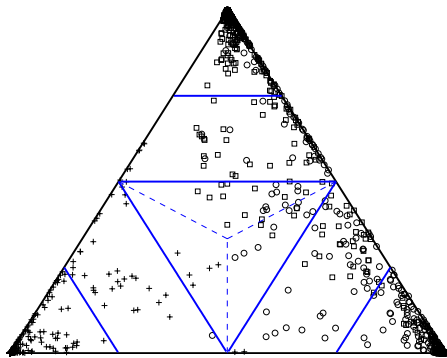


FIG. 6 – Probabilité pignistique avec l'opérateur de Dempster, $\square = \omega_1$, $\circ = \omega_2$, $+$ = ω_3

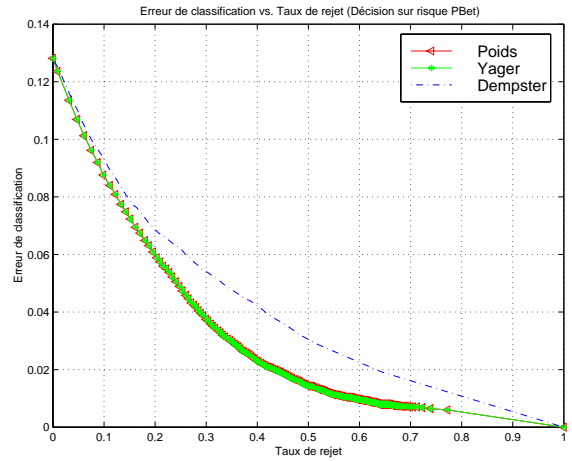


FIG. 7 – Taux d'erreur vs. taux de rejet pour le risque pignistique, $\cdot -$ = Dempster, $*$ = Yager, \triangleleft = Poids

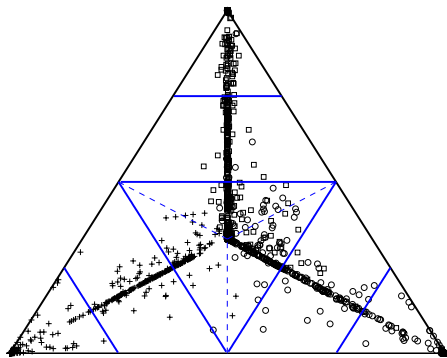


FIG. 8 – Probabilité pignistique avec l'opérateur de Yager, $\square = \omega_1$, $\circ = \omega_2$, $+$ = ω_3

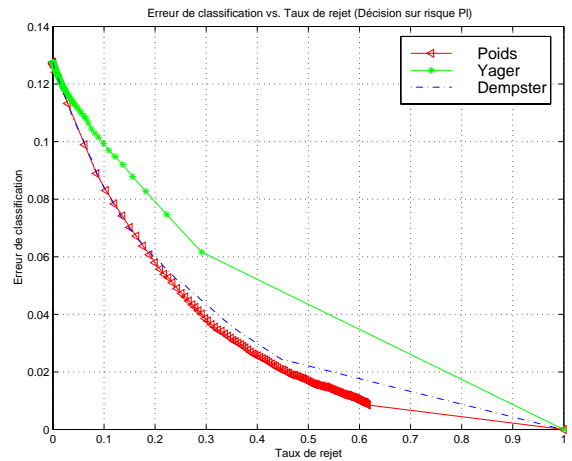


FIG. 9 – Taux d'erreur vs. taux de rejet pour le risque inférieur, $\cdot -$ = Dempster, $*$ = Yager, \triangleleft = Poids

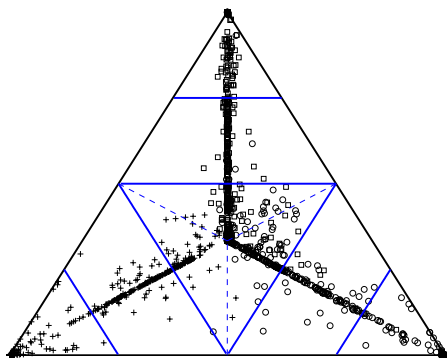


FIG. 10 – Probabilité pignistique avec l'opérateur pondéré, $\square = \omega_1$, $\circ = \omega_2$, $+$ = ω_3