

Estimation de méta-connaissances à partir de matrices de confusion pour la correction de fonctions de croyance

Estimation of meta-knowledge from confusion matrices to correct belief functions

E. Lefevre¹

F. Pichon¹

D. Mercier¹

Z. Elouedi²

B. Quost³

¹ Univ. Lille Nord de France, UArtois, EA 3926 LGI2A, France

{eric.lefevre,david.mercier,frederic.pichon}@univ-artois.fr

² Univeristé de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisie

zied.elouedi@isg.rnu.tn

³ Université de Technologie de Compiègne, UMR CNRS 7253 Heudiasyc, France

benjamin.quost@hds.utc.fr

Résumé :

Dans cet article, nous nous intéressons à la correction de fonctions de croyance et plus particulièrement à l'estimation de méta-connaissances nécessaire à la mise en place de ces corrections. Les méta-connaissances permettent de caractériser l'état dans lequel la source se trouve (par exemple : pertinente ou non, sincère ou non). Avec ces méta-connaissances, on peut donc corriger la fonction de croyance initialement donnée par la source d'information. Notre application concerne la fusion de classifieurs. Ainsi dans notre cas, les méta-connaissances sont estimées à partir de matrices de confusion. Les fonctions de croyance corrigées sont des fonctions catégoriques définies à partir des décisions fournies par les classifieurs.

Mots-clés :

Fonctions de croyance, Corrections, Méta-connaissances, Fusion de classifieurs.

Abstract:

In this paper, we are interested in the correction of belief functions, and more particularly the estimation of meta-knowledge needed for making these corrections. The meta-knowledge allow us to define the state in which the source is (for example : relevant or not, truthful or not). With these meta-knowledge, we can thus correct belief function initially given by the information source. Our application concerns the fusion of classifiers. So in our case, the meta-knowledge are estimated from confusion matrices. The corrected belief functions are categorical functions defined from the decisions induced from classifiers.

Keywords:

Belief functions, Corrections, Meta-knowledge, Fusion of classifiers.

1 Introduction

La théorie des fonctions de croyance constitue un cadre de travail riche et flexible pour la représentation et la manipulation des in-

formations imprécises et incertaines [10, 11]. Dans ce cadre, plusieurs mécanismes de correction des fonctions de masses [6, 9] sont disponibles : affaiblissement, renforcement,... Ces outils nécessitent une connaissance sur l'état de la source. Cette information est appelée *méta-connaissance* et caractérise l'état de la source : sincère ou non, pertinente ou non. Ces méta-connaissances sont généralement incertaines et par conséquent représentées sous forme de fonctions de croyance.

Dans ce papier, nous nous intéressons à l'estimation de méta-connaissances, à partir de matrices de confusion, pour la correction de fonctions de croyance dans le cadre de la fusion de classifieurs. Pour chacun de ces classifieurs, une matrice de confusion est disponible. Ces classifieurs fournissent également une décision qui est ensuite représentée sous forme d'une fonction de croyance. L'obtention de la matrice de confusion et l'apprentissage du classifieur sont réalisés sur deux bases de données distinctes rendant la méta-information et la décision du classifieur indépendantes. Dans cet article, nous proposons une approche pour estimer une méta-connaissance à partir de cette matrice de confusion afin de corriger la fonction de masse relative à la décision fournie par le classifieur.

Ce papier est organisé de la manière suivante.

Après avoir rappelé les concepts de la théorie des fonctions de croyance (Section 2), nous présentons un cadre général de correction introduit récemment dans la littérature [9] (Section 3). Les différentes méta-connaissances, utilisées pour la correction, et leurs estimations sont introduites dans la Section 4. Dans ce paragraphe, nous démontrons également que ces différentes approches peuvent être regroupées en deux modèles. La Section 5 permet d'étudier et de comparer ces modèles dans le cadre de la fusion de classifieurs sur des bases de données connues de la littérature.

2 Théorie des fonctions de croyance

2.1 Représentation de l'information

Le cadre de discernement $\mathcal{X} = \{\omega_1, \omega_2, \dots, \omega_K\}$ regroupe l'ensemble de toutes les propositions élémentaires ω_i relatives au problème à traiter.

L'impact d'un élément d'évidence sur les propositions de \mathcal{X} est représenté par une fonction appelée fonction de masse de croyance $m^{\mathcal{X}} : 2^{\mathcal{X}} \rightarrow [0; 1]$ telle que $\sum_{A \subseteq \mathcal{X}} m^{\mathcal{X}}(A) = 1$. Cette fonction affecte à chaque sous-ensemble $A \in \mathcal{X}$, un nombre $m(A) \in [0; 1]$ indiquant la croyance partielle assignée exactement à cette proposition. Les sous-ensembles de \mathcal{X} ayant une masse de croyance non nulle sont appelés les éléments focaux de $m^{\mathcal{X}}$. La fonction de masse de croyance vide telle que $m^{\mathcal{X}}(\mathcal{X}) = 1$ permet de représenter l'ignorance totale (cette fonction sera notée $m_{\mathcal{X}}^{\mathcal{X}}$ par la suite). Une fonction de masse de croyance est dite catégorique lorsqu'elle ne possède qu'un seul élément focal.

2.2 Combinaison

Soient $m_1^{\mathcal{X}}$ et $m_2^{\mathcal{X}}$, deux fonctions de masses de croyance définies sur le même cadre de discernement et issues de deux sources d'informations distinctes. Si les deux sources sont fiables, on peut appliquer la règle de combinai-

son conjonctive [11] définie par $\forall A \subseteq \mathcal{X}$:

$$m_1^{\mathcal{X}} \odot m_2^{\mathcal{X}}(A) = \sum_{B \cap C = A} m_1^{\mathcal{X}}(B) m_2^{\mathcal{X}}(C). \quad (1)$$

2.3 Décision

Pour la prise de décision, Smets [11] propose de transformer la fonction de masse de croyance en probabilité. Cette opération est appelée transformation pignistique et elle est définie par :

$$BetP(A) = \sum_{B \subseteq \mathcal{X}} \frac{|A \cap B|}{|B|} \frac{m^{\mathcal{X}}(B)}{(1 - m^{\mathcal{X}}(\emptyset))}, \quad \forall A \in \mathcal{X}. \quad (2)$$

2.4 Affaiblissement

Si l'on dispose d'une méta-connaissance sur la fiabilité de la source ayant fourni la fonction $m^{\mathcal{X}}$, il est possible de prendre en compte cette information par l'intermédiaire de l'opération d'affaiblissement [10, page 252]. L'affaiblissement de la fonction de masse $m^{\mathcal{X}}$, par une constante $\alpha \in [0, 1]$ appelé *taux d'affaiblissement*, est défini comme suit :

$$\begin{cases} \alpha m^{\mathcal{X}}(A) &= (1 - \alpha) m^{\mathcal{X}}(A), \quad \forall A \subset \mathcal{X}, \\ \alpha m^{\mathcal{X}}(\mathcal{X}) &= (1 - \alpha) m^{\mathcal{X}}(\mathcal{X}) + \alpha. \end{cases} \quad (3)$$

Lorsque la source d'information est totalement fiable ($\alpha = 0$), l'affaiblissement de la fonction de masse est sans effet; dans le cas contraire ($\alpha = 1$), il en résulte la fonction de masse vide. Un mécanisme de correction, généralisant cet affaiblissement, est présenté dans la section suivante.

3 Mécanisme de correction de fonctions de croyance

Dans [9], une approche générale permettant de prendre en compte des méta-connaissances sur les sources d'information a été proposée. Dans cette approche, on considère N comportements élémentaires $\mathcal{H} = \{h_1, \dots, h_N\}$ qui décrivent les états dans lesquels la source peut se trouver. On peut, par exemple, considérer

$\mathcal{H} = \mathcal{R} \times \mathcal{T}$ où $\mathcal{R} = \{R, \neg R\}$ représente la pertinence et $\mathcal{T} = \{T, \neg T\}$ la sincérité d'une source. Nous avons alors 4 comportements possibles pour la source dans l'ensemble $\mathcal{H} = \{(\mathcal{R}, \mathcal{T}), (\mathcal{R}, \neg \mathcal{T}), (\neg \mathcal{R}, \mathcal{T}), (\neg \mathcal{R}, \neg \mathcal{T})\}$. Si la source est supposée être dans l'un de ces états et que cet état influence les informations qu'elle fournit, il est nécessaire de les modifier en tenant compte des méta-connaissances disponibles. D'un point de vue formel, cette modification est réalisée par une mise en correspondance Γ_A entre la connaissance initiale $A \subseteq \mathcal{X}$ de la source et la nouvelle information produite en fonction de l'état dans lequel évolue la source. Ainsi $\Gamma_A(h)$ permet d'interpréter la connaissance $A \subseteq \mathcal{X}$ lorsque la source est dans l'état $h \in \mathcal{H}$. En reprenant, l'exemple précédent nous obtenons :

$$\begin{aligned}\Gamma_A(R, T) &= A \\ \Gamma_A(R, \neg T) &= \bar{A} \\ \Gamma_A(\neg R, T) &= \mathcal{X} \\ \Gamma_A(\neg R, \neg T) &= \mathcal{X}\end{aligned}$$

De façon plus générale, les méta-connaissances peuvent être incertaines et donc représentées par des fonctions de croyance $m^{\mathcal{H}}$ sur \mathcal{H} avec $\sum_{H \subseteq \mathcal{H}} m^{\mathcal{H}}(H) = 1$. De la même manière, les connaissances d'une source d'information S peuvent être représentées par une fonction de croyance $m_S^{\mathcal{X}}$ sur \mathcal{X} . Ainsi, sachant le comportement $H \subseteq \mathcal{H}$ de la source, chaque masse $m_S^{\mathcal{X}}(A)$ sera transférée à $\Gamma_A(H)$. La fonction de masse obtenue sera alors de la forme suivante :

$$m^{\mathcal{X}}(B) = \sum_H m^{\mathcal{H}}(H) \sum_{A: \Gamma_A(H)=B} m_S^{\mathcal{X}}(A). \quad (4)$$

Par exemple, si l'on considère une source totalement fiable mais non sincère, nous avons alors $m^{\mathcal{H}}(\{R, \neg T\}) = 1$. En appliquant l'équation (4), nous obtenons :

$$m^{\mathcal{X}}(\bar{A}) = m_S^{\mathcal{X}}(A) \quad \forall A \subseteq \mathcal{X}.$$

Cet exemple correspond à la négation d'une fonction de masse $\overline{m_S^{\mathcal{X}}}$ définie par : $\overline{m_S^{\mathcal{X}}}(A) = m_S^{\mathcal{X}}(\bar{A}) \forall A \subseteq \mathcal{X}$. Le mécanisme, ainsi défini,

est appelé correction fondée sur le comportement (*BBC*)¹ [9]. Il permet de généraliser des notions classiques de la théorie des fonctions de croyance telles que : l'affaiblissement [10] et le complémentaire d'une fonction de masse [2] (comme le démontre l'exemple précédent). Il existe un autre modèle de correction [6], appelé correction paramétrée (*PC*), qui est équivalent à *BBC* lorsque $m_S^{\mathcal{X}}$ est catégorique, ce qui est le cas considéré dans l'article.

4 Estimation des méta-connaissances nécessaires à la correction des fonctions de croyance

Dans cet article, la correction des fonctions de croyance est illustrée dans le cadre de la fusion de classifieurs. Les décisions de ces classifieurs seront transformées en masse de croyance catégorique ($m_S^{\mathcal{X}}(\{\omega_i\}) = 1$). En ce qui concerne les méta-connaissances, permettant de caractériser les comportements d'un classifieur, elles sont issues d'une matrice de confusion sans décision de rejet.

Une *matrice de confusion* $\mathcal{M} = (n_{k\ell})_{k \in \{1, \dots, K\}, \ell \in \{1, \dots, K\}}$ associée à une source est un tableau de contingence caractérisant les performances de cette source sur un ensemble de test (Tableau 1). Chaque ligne k correspond à une décision en faveur de ω_k . Chaque colonne ℓ correspond au cas où la vérité est ω_ℓ . Le terme général $n_{k\ell}$ est égal au nombre d'objets de classe ω_ℓ ayant été attribués à la classe ω_k par la source.

Nous noterons $n_{k\cdot} = \sum_{\ell=1}^K n_{k\ell}$ le nombre d'objets classés dans ω_k , où $k \in \{1, \dots, K\}$, et $n_{\cdot\ell} = \sum_{k=1}^K n_{k\ell}$ le nombre d'objets dont la classe réelle est ω_ℓ , où $\ell \in \{1, \dots, K\}$. Le nombre total d'objets classés dans la matrice est donné par $n = \sum_{k=1}^K \sum_{\ell=1}^K n_{k\ell}$.

Le *taux de classification correcte* \mathcal{T}_f d'une matrice est alors défini par sa trace $\mathcal{T}_f = \frac{\sum_{k=1}^K n_{kk}}{n}$.

1. *BBC* pour Behavior-Based Correction en anglais.

Tableau 1 – Illustration d’une matrice de confusion.

Décision \ Vérité	ω_1	\dots	ω_K
ω_1	n_{11}	\dots	n_{1K}
\vdots	\vdots	\ddots	\vdots
ω_K	n_{K1}	\dots	n_{KK}

On notera $\epsilon = 1 - \mathcal{T}_f$ le *taux d’erreur de classification*.

A la suite de cette section, différentes approches permettant d’évaluer le comportement d’une source à partir de sa matrice de confusion sont présentées. Toutes ces approches (déjà développées par ailleurs ou proposées ici) permettent d’évaluer la pertinence ou la sincérité de la source. Par conséquent, elles peuvent être considérées comme des approches définissant la fonction de masse $m^{\mathcal{H}}$ caractérisant la méta-connaissance sur le comportement de la source.

4.1 Pertinence d’une source à partir d’une matrice de confusion

Une première approche [3] consiste à considérer le taux de classification correcte comme son degré de pertinence :

$$\begin{aligned} m^{\mathcal{R}}(\{R\}) &= 1 - \epsilon \\ m^{\mathcal{R}}(\{R, \neg R\}) &= \epsilon. \end{aligned} \quad (5)$$

Il est à noter également qu’un affaiblissement, avec un coefficient de ϵ , est obtenu en considérant la fonction de masse suivante :

$$\begin{aligned} m^{\mathcal{R}}(\{R\}) &= 1 - \epsilon \\ m^{\mathcal{R}}(\neg\{R\}) &= \epsilon. \end{aligned} \quad (6)$$

4.2 Sincérité d’une source à partir d’une matrice de confusion

Une autre solution intéressante, consiste non pas à estimer la pertinence d’une source au travers de la matrice de confusion mais d’estimer

sa sincérité toujours à l’aide de la matrice de confusion tout en supposant que la source soit fiable. Dans ce cas, les prédictions mauvaises et les prédictions correctes peuvent être assimilées, respectivement, à des réalisations de la non-sincérité et de la sincérité de la source. En effet, on peut supposer que dans un problème de classification binaire, une source pertinente ne dit pas nécessairement la vérité si elle fait une erreur et dit la vérité si le résultat est correct. Une telle approche peut être exploitée au moins de trois façons différentes :

- en générant une fonction de masse bayésienne sur la sincérité d’une source :

$$\begin{aligned} m^{\mathcal{T}}(\{T\}) &= 1 - \epsilon \\ m^{\mathcal{T}}(\{\neg T\}) &= \epsilon. \end{aligned} \quad (7)$$

- en générant la fonction de masse issue du pari $BetP^{\mathcal{T}}(T) = 1 - \epsilon$ et $BetP^{\mathcal{T}}(\neg T) = \epsilon$. Ce pari correspond à la fonction de masse sur la sincérité de source donnée par :

$$\begin{aligned} m^{\mathcal{T}}(\{T\}) &= \begin{cases} 1 - 2\epsilon & \text{if } \epsilon \leq 1/2 \\ 0 & \text{ailleurs ;} \end{cases} \\ m^{\mathcal{T}}(\{\neg T\}) &= \begin{cases} 0 & \text{if } \epsilon \leq 1/2 \\ 2\epsilon - 1 & \text{ailleurs ;} \end{cases} \\ m^{\mathcal{T}}(\mathcal{T}) &= \begin{cases} 2\epsilon & \text{if } \epsilon \leq 1/2 \\ 2(1 - \epsilon) & \text{ailleurs ;} \end{cases} \end{aligned} \quad (8)$$

- en générant une fonction de masse sur la sincérité de la source en utilisant les travaux de Dempster [1] :

$$\begin{aligned} m^{\mathcal{T}}(\{T\}) &= \frac{1 - \epsilon}{1 + 1/n} \\ m^{\mathcal{T}}(\{\neg T\}) &= \frac{\epsilon}{1 + 1/n} \\ m^{\mathcal{T}}(\mathcal{T}) &= \frac{1}{n + 1} \end{aligned} \quad (9)$$

4.3 Discussion

L’influence des méta-connaissances sur le comportement des sources d’information peuvent s’exprimer de manière uniforme. En effet, en reprenant la solution proposée à l’équation (5), la

fonction de masse obtenue sur \mathcal{X} , en appliquant *BBC*, peut s'exprimer par :

$$m^{\mathcal{X}} = (1 - \epsilon)m_S^{\mathcal{X}} + \epsilon m_{\mathcal{X}}^{\mathcal{X}}. \quad (10)$$

De la même manière, la fonction de masse issue de la méta-connaissance exprimée dans l'équation (7) devient :

$$m^{\mathcal{X}} = (1 - \epsilon)m_S^{\mathcal{X}} + \overline{\epsilon m_S^{\mathcal{X}}}. \quad (11)$$

Nous pouvons noter également que cette dernière expression est équivalente à celle proposée par Xu et al. [12] dans le cas où $m_S^{\mathcal{X}}$ est catégorique. La fonction de masse obtenue avec les connaissances par l'intermédiaire de l'équation (8) est donnée par (en considérant $\epsilon \leq 0.5$) :

$$m^{\mathcal{X}} = (1 - 2\epsilon)m_S^{\mathcal{X}} + 2\epsilon m_{\mathcal{X}}^{\mathcal{X}}. \quad (12)$$

La fonction de masse obtenue à partir de la méta-connaissance exprimée par l'équation (9) est définie par :

$$m^{\mathcal{X}} = \frac{1 - \epsilon}{1 + 1/n} m_S^{\mathcal{X}} + \frac{\epsilon}{1 + 1/n} \overline{m_S^{\mathcal{X}}} + \frac{1}{n + 1} m_{\mathcal{X}}^{\mathcal{X}} \quad (13)$$

A partir de ces équations, nous pouvons faire plusieurs remarques. Dans un premier temps, les expressions du modèle de pertinence classique (5) et du modèle bayésien sur la véracité (7) diffèrent uniquement par leur dernier membre ($m_{\mathcal{X}}^{\mathcal{X}}$ au lieu de $\overline{m_S^{\mathcal{X}}}$). Le modèle proposé par l'équation (8) et le modèle de pertinence classique (5) font une combinaison linéaire des mêmes fonctions de masses à un coefficient près. Seul le dernier modèle (9) caractérisant la sincérité d'une source semble différent des autres. Toutefois, lorsque n est grand, ce modèle est proche du modèle bayésien (11). Ainsi, uniquement deux modèles apparaissent :

$$m^{\mathcal{X}} = (1 - \epsilon)m_S^{\mathcal{X}} + \overline{\epsilon m_S^{\mathcal{X}}} \quad (14)$$

et

$$m^{\mathcal{X}} = (1 - \alpha\epsilon)m_S^{\mathcal{X}} + \alpha\epsilon m_{\mathcal{X}}^{\mathcal{X}} \quad (15)$$

avec $\alpha = 1, 2$. Ces différents modèles sont étudiés et comparés dans la section suivante.

5 Expérimentation

5.1 Méthodologie :

Soit une base de données B constituée de n instances. On obtient les résultats des différents classifieurs de la manière suivante :

1. De la base B , on retire une instance qui constituera à elle seule la base de test,
2. Le reste des instances de B sont réparties en deux bases de taille équivalente ($\frac{n-1}{2}$) : B_{train} et B_{conf} ,
3. Les différents classifieurs utilisés sont entraînés sur la base B_{train} ,
4. Les instances de la base B_{conf} sont testées afin d'obtenir les matrices de confusion de chaque classifieur.
5. A partir de ces matrices de confusion, on calcule pour chaque classifieur le taux de classification correcte \mathcal{T}_f . Par ailleurs, afin de comparer les approches d'estimation que nous proposons dans cet article, nous reprenons des mécanismes de correction contextuels définis dans [3]. Pour cela deux taux de fiabilité contextuels, $\mathcal{T}_f[\cdot]$ et $\mathcal{T}_{fd}[\cdot]$, sont calculés.
6. On teste ensuite l'instance extraite à l'étape 1 à l'aide de chaque classifieur. La sortie des classifieurs est alors considérée comme une fonction de masse catégorique.
7. On corrige ces fonctions de masses en considérant les taux de classification correcte comme les estimations des méta-connaissances et en les intégrant dans les différentes approches proposées dans la section 4.3. On appelle le modèle donné par l'équation (10) *Mod. 1*, le modèle de l'équation (11) *Mod. 2* et le modèle de l'équation (12) *Mod. 3*. Par ailleurs, une autre stratégie consiste à utiliser les deux coefficients contextuels, calculés à l'étape 5, pour corriger les fonctions de masses catégoriques [3, 7]. Ces deux approches seront nommées $\mathcal{T}_f[\cdot]$ et $\mathcal{T}_{fd}[\cdot]$ par la suite.

8. On fusionne les masses de croyance corrigées des classifieurs à l'aide de la règle conjonctive.
9. On emploie la règle du maximum de probabilité pignitistique afin de classer l'instance testée.

Ces différentes opérations sont répétées autant de fois qu'il y a de vecteurs dans la base B . Par ailleurs, pour chaque instance testée, il y a 20 tirages aléatoires afin de constituer B_{train} et B_{conf} . Les résultats présentés sont les moyennes de ces 20 tirages.

5.2 Mesures de performance :

Afin de mesurer les performances des différentes méthodes, on calcule :

- le taux de classification correcte moyen et l'écart-type moyen,
- la F-mesure (et l'écart type associé) qui permet de combiner la précision et le rappel. La mesure F_{β}^k d'une classe k pour une classe k est donnée par :

$$F_{\beta}^k = \frac{(1 + \beta^2)precision^k.rappel^k}{\beta^2.precision^k + rappel^k} \quad (16)$$

avec :

$$precision^k = \frac{n_{kk}}{n_{k.}} \quad \text{et} \quad rappel^k = \frac{n_{kk}}{n_{.k}}$$

La mesure F_{β} globale du classifieur est obtenue en faisant la moyenne sur l'ensemble des classes :

$$F_{\beta} = \frac{\sum_{k=1}^K F_{\beta}^k}{K}. \quad (17)$$

Parmi les F_{β} mesures, la plus utilisée est la F_1 ($\beta = 1$). Dans ce cas, cette mesure correspond à la moyenne, pour toutes les classes, des moyennes harmoniques entre la précision et le rappel.

Les résultats des approches proposées sont comparés aux résultats obtenus avec le vote majoritaire et aux performances des classifieurs pris individuellement.

5.3 Classifieurs et bases de données employées :

Actuellement, trois classifieurs ont été utilisés : le 1 plus proche voisin (1-NN), les Réseaux Bayésiens Naïfs (NB) et les Arbres de décision (DT) [4]. Les approches ont été testées sur différentes bases de données réelles issues de la base U.C.I. [8]. Une description rapide de ces bases est présentée dans le tableau 2.

Tableau 2 – Description des bases de données.

Database	Ref	#instances	#attributes
Balance Scale	BS	625	4
Breast Cancer	BC	286	9
Car	CA	1728	6
Diabetes	DI	768	8
Ionosphere	IO	351	34
Iris	IR	150	4
Segment Challenge	SC	1500	19
Wine	WI	178	13
Zoo	ZO	101	17

5.4 Résultats et commentaires

Les résultats obtenus, en terme de taux de classification correcte, sont présentés dans les tableaux 3 et 4.

A la lecture de ces résultats, on constate dès lors que l'un des classifieurs a un taux de classification correcte plus faible que les deux autres, les approches proposées donnent des résultats meilleurs (ou similaires) à ceux obtenus avec le vote majoritaire. C'est le cas, par exemple, pour les bases de données BS , CA , SC , WI et ZO .

Toutefois, les résultats obtenus avec les approches proposées ne sont pas significativement meilleurs (voire parfois moins bons) que le meilleur des classifieurs. Il est à noter que chaque classifieur est optimisé lors de la phase d'apprentissage (sur la base B_{Train}). Ainsi, les performances des classifieurs individuels sont élevées. Cela explique peut-être les écarts de performance relativement faibles entre le

Tableau 4 – Taux de classification correcte et écart-type avec les différentes approches de fusion de classifieurs.

REF	Vote Maj.	Mod. 1	Mod. 2	Mod. 3	$\mathcal{T}_f[\cdot]$	$\mathcal{T}_{fa}[\cdot]$
BS	0.8479 (0.0072)	0.8644 (0.0077)	0.8650 (0.0079)	0.8695 (0.0072)	0.5908 (0.0140)	0.7375 (0.0131)
BC	0.7336 (0.0115)	0.7336 (0.0115)	0.7336 (0.0115)	0.7336 (0.0115)	0.6969 (0.0169)	0.6969 (0.0169)
CA	0.8389 (0.0050)	0.8600 (0.0063)	0.8600 (0.0063)	0.8602 0.0062	0.8144 (0.0042)	0.8203 (0.0044)
DI	0.7527 (0.0119)	0.7527 (0.0119)	0.7527 (0.0119)	0.7526 (0.0119)	0.7421 (0.0130)	0.7421 (0.0130)
IO	0.9175 (0.0062)	0.9175 (0.0062)	0.9173 (0.0064)	0.9170 (0.0062)	0.9081 (0.0109)	0.9081 (0.0109)
IR	0.9500 (0.0070)	0.9500 (0.0070)	0.9500 (0.0070)	0.9500 (0.0070)	0.9447 (0.0072)	0.9447 (0.0072)
SC	0.9447 (0.0036)	0.9493 (0.0040)	0.9493 (0.0040)	0.9493 (0.0040)	0.9477 (0.0027)	0.9478 (0.0032)
WI	0.9615 (0.0080)	0.9640 (0.0072)	0.9640 (0.0072)	0.9682 (0.0080)	0.9653 (0.0091)	0.9653 (0.0091)
ZO	0.9421 (0.0188)	0.9460 (0.0196)	0.9460 (0.0196)	0.9460 (0.0196)	0.9302 (0.0168)	0.9302 (0.0168)

Tableau 3 – Taux de classification correcte et écart-type avec les classifieurs 1-NN, NB et DT.

REF	DT	1-NN	NB
BS (288/49/288)	0.7832 (0.0100)	0.7712 (0.0093)	0.8895 (0.0047)
BC (201/85) (201/85)	0.7154 (0.0160)	0.7187 (0.0162)	0.7257 (0.0115)
CA (1210/384/69/65)	0.8816 (0.0042)	0.7136 (0.0021)	0.8418 (0.0047)
DI (500/268)	0.7280 (0.0115)	0.7022 (0.0157)	0.7496 (0.0051)
IO (126/225)	0.8836 (0.0116)	0.8570 (0.0106)	0.8252 (0.0077)
IR (50/50/50)	0.9337 (0.0930)	0.9467 (0.0099)	0.9487 (0.0061)
SC (205/220/208/220/204/236/207)	0.9419 (0.0048)	0.9470 (0.0034)	0.8089 (0.0033)
WI (59/71/48)	0.8907 (0.0179)	0.9471 (0.0097)	0.9710 (0.0080)
ZO (41/20/5/13/4/8/10)	0.8767 (0.0180)	0.9450 (0.0138)	0.9317 (0.0176)

meilleur classifieur et les approches de fusion.

On constate également que les résultats obtenus avec les approches fondées sur un af-

faiblissement contextuel sont moins bons que ceux obtenus à l'aide des approches basées sur des corrections non contextuelles et étudiées dans cet article (*Mod.1* à *3*). Cette remarque est d'autant plus vraie lorsqu'il y a une disparité de représentation importante entre les classes (comme c'est le cas par exemple de la base de données *BS*)². Dans ce cas, les classifieurs simples et les corrections non contextuelles ont tendance à négliger les classes sous représentées ce qui n'est pas le cas des méthodes contextuelles. On peut constater ce phénomène en regardant les valeurs de la F-mesure (Figures 1 et 2) sur les bases *BS* et *CA*. Les figures montrent que pour $\beta = 1$, les approches contextuelles ont de meilleurs résultats que les autres approches (ou très proches) au contraire des résultats obtenus dans le tableau 4.

6 Conclusion

Dans cet article, le problème de l'estimation de méta-connaissances, à partir de matrices de

2. Les valeurs entre parenthèses, en dessous du nom de la base dans le Tableau 3, indiquent le nombre d'instances de chaque classe.

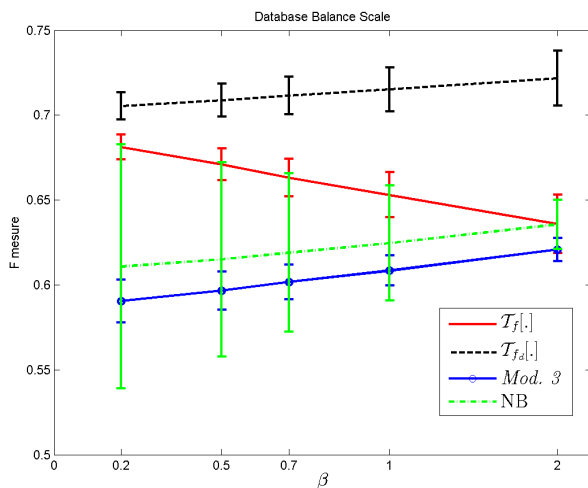


Figure 1 – Evolution de la F_β mesure en fonction de β sur la base de données BS.

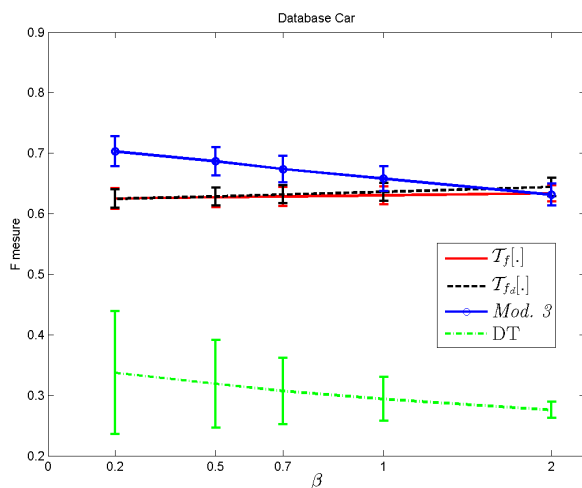


Figure 2 – Evolution de la F_β mesure en fonction de β sur la base de données CA.

confusion, pour la correction de fonctions de croyance a été abordé. Il a été démontré que les estimations de méta-connaissances pouvaient être regroupées sous la forme de deux modèles. Une expérimentation dans le cadre de la fusion de classifieurs a été menée. Celle-ci a montré que la fusion de fonctions de croyance corrigées à l'aide de ces méta-connaissances permettait d'obtenir des résultats meilleurs (ou similaires) au vote majoritaire.

En terme de perspectives, une étude sur l'estimation des méta-connaissances à l'aide de matrices de confusion contenant l'hypothèse de rejet est envisagée. Par ailleurs, une application pour la fusion de classifieurs crédaux (au lieu de classifieurs crips comme c'est le cas dans notre étude) permettrait d'utiliser d'autres mécanismes de correction comme le renforcement [6].

Références

- [1] A.P. Dempster. New methods of reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, 37 :355-374, 1966.
- [2] D. Dubois et H. Prade. A set-theoretic view of belief functions : logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12(3) :193-226, 1986.
- [3] Z. Elouedi, E. Lefevre, et D. Mercier. Discountings of a Belief Function Using a Confusion Matrix, in *Proceedings of the 22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 2010*, vol. 1, pp. 287-294.
- [4] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann et I. H. Witten. The WEKA Data Mining Software : An Update. *SIGKDD Explorations*, 11 (1), 2009.
- [5] D. Mercier, B. Quost et T. Denœux. Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextual Discounting. *Information Fusion*, 9(2) :246-258, 2008.
- [6] D. Mercier, T. Denœux et M.-H. Masson, Belief function correction mechanisms, in : B. Bouchon-Meunier, R.R. Yager, J.-L. Verdegay, M. Ojeda-Aciego, L. Magdalena (Eds.), *Foundation of Reasoning under Uncertainty, Studies in Fuzziness and Soft Computing*, vol. 249, Springer Verlag, Berlin, pp. 203-222, 2010.
- [7] D. Mercier. Extending the contextual discounting of a belief function thanks to its canonical disjunctive decomposition. *1st Workshop on the Theory of Belief Functions (BELIEF), 2010*, Brest, France, paper 61.
- [8] MP. Murphy et D. W. Aha. UCI repository databases. <http://www.ics.uci.edu/mllearn>, 1996.
- [9] F. Pichon, D. Dubois et T. Denœux. Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53(2) :159-175, 2012.
- [10] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.
- [11] Ph. Smets et R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66 :191-243, 1994.
- [12] L. Xu, A. Krzyzak et C.Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22 (3) :418-435, 1992.