

Corrections contextuelles crédibilistes en fonction de décisions partielles

S. Mutmainah^{1,2}

S. Hachour¹

F. Pichon¹

D. Mercier¹

¹ Univ. Artois, EA 3926 LGI2A, Béthune, F-62400, France.

² UIN Sunan Kalijaga, Yogyakarta, Indonesia.

siti.mutmainah@uin-suka.ac.id

samir.hachour@univ-artois.fr

frederic.pichon@univ-artois.fr

david.mercier@univ-artois.fr

Résumé :

Dans ce papier, une amélioration de la qualité d'une source d'information crédibiliste est proposée en utilisant des corrections contextuelles dépendant de décisions partielles obtenues à partir d'une relation de dominance forte sur les sorties de la source. Des expériences numériques avec le classifieur EkNN et des données synthétiques et réelles permettent d'illustrer les performances et l'intérêt de cette méthode.

Mots-clés :

Fonctions de croyance, Corrections contextuelles, Décisions partielles, Dominance forte.

Abstract:

In this paper, an improvement of the quality of an evidential source of information is proposed using contextual corrections depending on partial decisions obtained from an interval dominance relation on the source outputs. Numerical experiments with the EkNN classifier and synthetic and real data allow one to illustrate the performances and the interest of this method.

Keywords:

Belief functions, Contextual corrections, Partial decisions, Interval dominance.

1 Introduction

En reconnaissance des formes [1], la qualité des informations fournies par une source (*par exemple* un capteur, un classifieur, . . .) joue un rôle important dans la réussite de l'opération de reconnaissance. En effet, les informations fournies par une source peuvent être fausses, biaisées ou non pertinentes.

La théorie des fonctions de croyance de Demspter-Shafer [16, 18] fournit un cadre mathématique flexible pour représenter et manipuler ces informations imparfaites. Dans cette théorie, la qualité d'une source d'information

est classiquement traitée à l'aide de l'opération d'*affaiblissement* (ou *discounting* en anglais) introduite par G. Shafer dans son livre [16, chapitre 11, page 251]. Cette méthode a depuis été affinée grâce à des mécanismes de correction dits contextuels [11, 15] prenant en compte des connaissances plus fines sur la qualité d'une source : sa fiabilité (au sens de sa pertinence, c'est-à-dire la capacité de la source à répondre à la question d'intérêt) et sa sincérité (c'est-à-dire sa capacité à dire ce qu'elle sait ; cette capacité pouvant être consciente - comme un mensonge par exemple - ou inconsciente - comme un biais par exemple) [14]. Plus précisément, trois mécanismes ont été introduits. Ils sont respectivement appelés *affaiblissement contextuel* (ou *Contextual Discounting (CD)*), *renforcement contextuel* (ou *Contextual Reinforcement (CR)*) et *reniement contextuel* (ou *Contextual Negating (CN)*). Ils peuvent tous être mathématiquement dérivés de ces notions de fiabilité et de sincérité [15]. La correction CD, qui généralise l'opération d'affaiblissement, peut ajuster la sortie d'une source en fonction des informations sur sa fiabilité, tandis que la correction CN, qui généralise le reniement (*negation* en anglais) d'une source [8], peut ajuster la source en fonction de sa sincérité, enfin, la correction CR est l'opération duale de CD, elle peut renforcer une source trop prudente [13, 15].

Dans ce papier, nous proposons d'améliorer la qualité d'une source d'information fournissant

Tableau 1 – Exemple d’une donnée étiquetée considérée : la source a donné une information m_S concernant la vraie classe d’un objet o appartenant à $\Omega = \{\omega_1, \omega_2, \omega_3\}$

	ω_1	ω_2	ω_3	ω_1	ω_1	ω_2	Ω	Vérité
				ω_2	ω_3	ω_3		
$m_S\{o\}$	0	0	0.5	0	0	0.3	0.2	ω_1

en sortie une fonction de croyance concernant la réponse à une question d’intérêt en utilisant des corrections contextuelles CD, CR ou CN, et des décisions partielles calculées à partir des sorties de cette source en utilisant la relation de dominance d’intervalle (*interval dominance*) [19, 20], également appelée dominance forte (*strong dominance*) dans [6, 10]. La source est considérée comme une *boîte noire*, nous n’avons pas accès à la manière dont elle fonctionne pour fournir ses sorties. Cette situation se produit par exemple lorsqu’une entreprise achète des capteurs ou des machines en charge de reconnaître des objets à d’autres entreprises possiblement concurrentes, et que le processus de prise de décision (ou l’algorithme) utilisé par ce capteur ou cette machine n’est pas connu, voire est protégé [12]. Un ensemble d’apprentissage est disponible. Il est composé de sorties de la source (fonctions de masse par exemple) concernant des données dont la vérité terrain est connue. Par exemple [9], on peut avoir dans l’ensemble d’apprentissage l’information suivante m_S fournie par la source concernant la vraie classe d’un objet o appartenant à $\Omega = \{\omega_1, \omega_2, \omega_3\}$ (cf Tableau 1).

Au lieu de choisir une meilleure correction possible entre CD, CR et CN à partir de l’ensemble d’apprentissage comme cela a été proposé dans [15], nous souhaitons dans ce travail proposer une nouvelle méthode qui tire profit de toutes les corrections, et proposons de regrouper les sorties conduisant aux mêmes décisions partielles pour apprendre la meilleure correction possible entre CD, CR et CN au sein de chacun de ces groupes. Avec cette stratégie, une sortie est donc corrigée différemment selon la

décision partielle à laquelle elle conduit.

Ce papier est organisé de la manière suivante. Dans la section 2, les concepts et notations sur les fonctions de croyance utilisés dans ce papier sont présentés, un rappel est également fait sur la prise de décision avec la relation de dominance forte utilisée avec les fonctions de croyance. Des rappels sur les corrections contextuelles sont donnés dans la section 3. Par la suite, dans la section 4, la méthode proposée pour apprendre les corrections contextuelles en fonction des décisions partielles est exposée. Elle est testée avec des données synthétiques et réelles dans la section 5, une discussion étant également ajoutée dans cette dernière section.

2 Fonctions de croyance : Concepts de base utilisés et notations

Dans cette section, nous rappelons rapidement les concepts et notations nécessaires utilisés dans ce papier (voir par exemple [16, 17, 4] pour plus de détails).

Nous noterons $\Omega = \{\omega_1, \dots, \omega_K\}$ l’univers (ou le cadre de discernement) représentant l’ensemble fini des réponses à une question d’intérêt donnée. Une information concernant la réponse à cette question d’intérêt induit une *fonction de masse* m^Ω (notée simplement m en l’absence d’ambiguïté) définie de 2^Ω dans $[0, 1]$, vérifiant $\sum_{A \subseteq \Omega} m^\Omega(A) = 1$. Les *éléments focaux* d’une fonction de masse m sont les sous-ensembles $A \subseteq \Omega$ t.q. $m(A) > 0$. Une fonction de masse n’ayant qu’un seul élément focal A est appelée *fonction de masse logique ou catégorique* et peut être simplement notée par m_A .

Une fonction de masse m est en bijection avec une *fonction de croyance* Bel et une *fonction de plausibilité* Pl , qui sont respectivement définies pour tout $A \subseteq \Omega$ par $Bel(A) = \sum_{B \subseteq A} m(B)$, et $Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$.

La *fonction de contour* pl correspond à la restriction de la fonction de plausibilité aux sin-

gletons de Ω , elle est définie pour tout $\omega \in \Omega$ par $pl(\omega) = Pl(\{\omega\})$.

Si une source S fournit une information m_S , et que l'on sait que cette source est fiable avec un degré de croyance $\beta = 1 - \alpha \in [0, 1]$, alors cette fonction de masse originale m_S peut être *affaiblie* en une FM m t.q.

$$m = \begin{cases} A & \mapsto \beta m_S(A) & \forall A \subset \Omega \\ \Omega & \mapsto \beta m_S(\Omega) + \alpha \end{cases} \quad (1)$$

L'équation précédente (1) peut aussi simplement s'écrire $m = \beta m_S + \alpha m_\Omega$, avec m_Ω la fonction de masse catégorique définie par $m_\Omega(\Omega) = 1$, et on peut aussi montrer (voir par exemple [15, Prop. 11]) que la fonction de contour associée à la fonction de masse affaiblie m est définie pour tout $\omega \in \Omega$ par $pl(\omega) = 1 - (1 - pl_S(\omega))\beta$, avec pl_S la fonction de contour de m_S . Des justifications formelles de cette opération peuvent être trouvées dans [17, 11, 15].

Enfin, lorsque le moment est venu de prendre une décision, considérons que l'ensemble des décisions ou des actes possibles est donné par $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$. Considérons également, que le cadre de discernement Ω représente les états de la nature, et que nous disposons d'une fonction d'utilité u définie de $\mathcal{A} \times \Omega$ dans \mathbb{R} qui indique par un nombre réel l'utilité de chaque acte $a \in \mathcal{A}$ lorsque la vérité est ω , pour tout $\omega \in \Omega$. Alors [3, 6] les utilités espérées inférieure et supérieure de tout acte a , la réalité connue étant m , peuvent être respectivement définies par

$$\underline{E}(a) = \sum_{B \subseteq \Omega} m(B) \min_{\omega \in B} u(a, \omega), \quad (2)$$

et

$$\overline{E}(a) = \sum_{B \subseteq \Omega} m(B) \max_{\omega \in B} u(a, \omega). \quad (3)$$

Une manière de prendre une décision [3, 6] consiste alors à considérer une relation de dominance [19, 10] en considérant un ensemble d'éléments non dominés selon une relation de préférence, qui peut être qu'un acte a est préféré

à un autre acte a' , si et seulement si son utilité espérée inférieure est plus grande que l'utilité espérée supérieure de a' (relation de dominance forte [6, 10]), formellement :

$$a \succeq a' \iff \underline{E}(a) \geq \overline{E}(a') \quad (4)$$

Si maintenant, on considère que l'ensemble des actes est égal à Ω , $a = \omega$ signifiant "la réponse à la question d'intérêt est ω ", et si on considère également des utilités 0 – 1, c'est-à-dire $u(\omega, \omega) = 1$ et $u(\omega, \omega') = 0$ pour tout $\omega \in \Omega, \omega' \in \Omega$ et $\omega \neq \omega'$, on peut alors voir que $\underline{E}(\omega) = Bel(\{\omega\})$ et $\overline{E}(\omega) = Pl(\{\omega\})$, pour tout $\omega \in \Omega$, et la relation (4) devient

$$\omega \succeq \omega' \iff Bel(\{\omega\}) \geq Pl(\{\omega'\}) \quad (5)$$

Étant donné une fonction de masse m , les *singletons non dominés* ω de m seront alors les singletons ω tels qu'il n'existe pas de $\omega' \in \Omega$ avec $Bel(\{\omega'\}) \geq Pl(\{\omega\})$.

3 Corrections contextuelles et apprentissage

Dans cette section, les définitions de CD, CR et CN sont rappelées ainsi que les possibilités de les apprendre à partir de données étiquetées [15] composées des sorties d'une source concernant des objets dont la vraie classe est connue.

Les fonctions de contour résultantes d'une application de CD, CR et CN sont rappelées ici dans le Tableau 2 dans le cas de K paramètres, où K est le nombre d'éléments de Ω .

Ces configurations sont en effet suffisamment expressives pour atteindre les valeurs minimales (avec ces mécanismes) de la mesure de divergence E_{pl} suivante [11, 15] entre les sorties corrigées par ces mécanismes et les vérités terrain de chaque objet dans l'ensemble d'apprentissage :

$$E_{pl}(\beta) = \sum_{i=1}^n \sum_{k=1}^K (pl_i(\omega_k) - \delta_{i,k})^2, \quad (6)$$

Tableau 2 – Fonctions de contour de chaque correction contextuelle d’une fonction de masse m_S données pour tout $\omega \in \Omega$. Chaque paramètre β_ω peut varier dans $[0, 1]$.

Corr.	Fonctions de contour
CD	$pl(\omega) = 1 - (1 - pl_S(\omega))\beta_\omega$
CR	$pl(\omega) = pl_S(\omega)\beta_\omega$
CN	$pl(\omega) = .5 + (pl_S(\omega) - .5)(2\beta_\omega - 1)$

où n est le nombre d’objets dans l’ensemble d’apprentissage, $\beta = (\beta_\omega, \omega \in \Omega)$ est le vecteur composé des K paramètres de chaque correction, pl_i est la fonction de contour au regard de la classe de l’objet o_i résultante d’une correction contextuelle (CD, CR ou CN) de la fonction de masse fournie par la source pour cet objet o_i , et $\delta_{i,k}$ est la fonction indicatrice de la vérité des objets $o_i, i \in \{1, \dots, n\}$ t.q. $\delta_{i,k} = 1$ si la classe de l’objet o_i est ω_k , sinon $\delta_{i,k} = 0$.

Trouver les K paramètres $\beta = (\beta_\omega, \omega \in \Omega)$ minimisant la mesure E_{pl} pour un ensemble d’apprentissage donné aboutit alors à un problème à moindre carré linéaire qui peut être résolu efficacement à l’aide d’algorithmes standard.

Par exemple, pour CD, E_{pl} peut s’écrire

$$E_{pl}(\beta) = \|\mathbf{Q}\beta - \mathbf{d}\|^2 \quad (7)$$

avec

$$\mathbf{Q} = \begin{bmatrix} \text{diag}(\mathbf{pl}_1 - 1) \\ \vdots \\ \text{diag}(\mathbf{pl}_n - 1) \end{bmatrix}, \mathbf{d} = \begin{bmatrix} \delta_1 - 1 \\ \vdots \\ \delta_n - 1 \end{bmatrix} \quad (8)$$

où $\text{diag}(\mathbf{v})$ est une matrice diagonale carrée dont la diagonale est composée des éléments du vecteur \mathbf{v} , et où pour tout $i \in \{1, \dots, n\}$, δ_i est le vecteur colonne composé des valeurs de la fonction indicatrice δ_i , soit $\delta_i = (\delta_{i,1}, \dots, \delta_{i,K})^T$.

4 Corrections contextuelles en fonction de décisions partielles

Dans ce papier, nous proposons d’améliorer la méthode d’apprentissage exposée dans la section 3 en considérant, dans l’ensemble d’apprentissage, des groupes de décisions partielles distinctes auxquelles peuvent conduire les sorties de la source. Les décisions partielles sont calculées en utilisant la relation (5) et sont composées de l’ensemble des singletons non dominés. Les meilleurs paramètres β pour chaque correction CD, CR et CN selon E_{pl} (6) sont alors calculés pour chaque groupe. La correction ayant la plus faible valeur de E_{pl} dans chaque groupe est alors conservée, c’est cette correction qui sera employée dans l’ensemble de test. Cette procédure d’apprentissage est résumée par l’Algorithme 1.

Algorithme 1 Procédure d’apprentissage

Entrée : Un ensemble \mathcal{I} d’instances $(m_i\{o_i\}, \omega_i)$, $i \in \{1, \dots, n\}$, $m_i\{o_i\}$ étant la sortie de la source concernant l’objet o_i dont la classe est ω_i .

Sortie : Groupes \mathcal{G} de décisions partielles avec la meilleure correction selon E_{pl} (6) pour chaque groupe.

- 1: **procédure** APPRENTISSAGE
 - 2: \mathcal{G} initialement vide
 - 3: **pour** chaque instance i dans \mathcal{I} **faire**
 - 4: Calculer la décision partielle associée à m_i selon la relation (5)
 - 5: Ajouter i au groupe g de \mathcal{G} associé à cette décision partielle.
 - 6: **pour** chaque groupe de décisions partielles g dans \mathcal{G} **faire**
 - 7: Calculer les meilleures paramètres de CD, CR et CN pour minimiser E_{pl} (6) selon les instances de ce groupe
 - 8: Garder la correction obtenant la plus petite valeur pour E_{pl} .
-

5 Applications numériques et discussion

Dans cette section, nous présentons des expériences réalisées avec le classifieur EkNN [2, 5] avec $k = 5$ comme source d'information sur des données synthétiques et réelles.

Les données synthétiques sont composées de 5000 instances, 5 classes et 2 attributs, qui ont été générées à partir d'une distribution normale multivariée avec des moyennes $\mu_1 = (0, 0)$, $\mu_2 = (2, 0)$, $\mu_3 = (0, 2)$, $\mu_4 = (2, 2)$, $\mu_5 = (1, 1)$ pour respectivement les classes 1, 2, 3, 4 et 5, et la même matrice de covariance Σ pour chaque classe :

$$\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \quad (9)$$

Une illustration d'un ensemble de données synthétiques générées est donnée dans la Figure 1. À partir de ces mêmes données, on peut voir sur la Figure 2 pour un certain nombre de points de l'espace des attributs des groupes de décisions partielles obtenus à partir des sorties de la source sur ces données. Il faut bien noter que ces groupes de décisions partielles ne sont pas nécessairement tous obtenus dans la phase d'apprentissage décrite dans l'Algorithme 1. Ces points sont simplement donnés ici à titre d'illustration.

Les données réelles utilisées sont issues de bases UCI [7] décrites dans le Tableau 3.

Pour chaque ensemble de données, l'expérience suivante est répétée 10 fois :

- Une moitié des données (\mathcal{L}_1) est utilisée pour apprendre le classifieur EkNN ;
- Une validation croisée à 10 blocs (10-fold cross validation) est ensuite réalisée sur la deuxième moitié des données avec 9 blocs (\mathcal{L}_2) pour apprendre la meilleure correction dans chaque groupe de décisions partielles en utilisant l'Algorithme 1, et 1 bloc pour la phase de test.

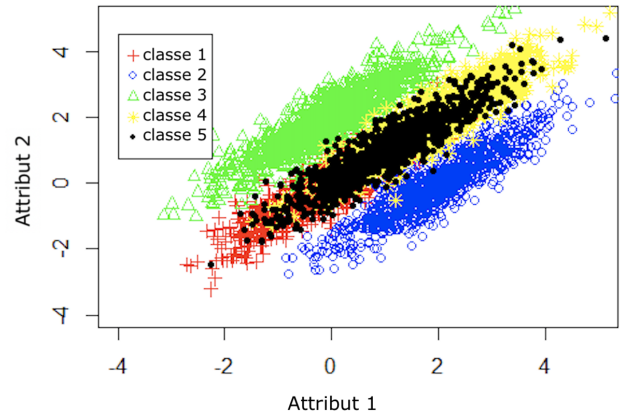


Figure 1 – Illustration d'un ensemble généré de données composées de 5 classes et de 2 attributs.

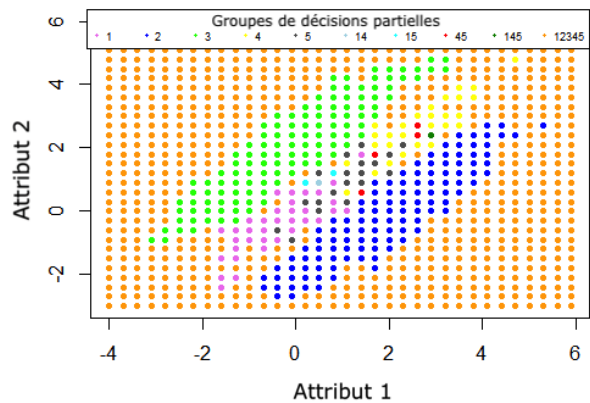


Figure 2 – Illustrations de l'appartenance de certains points de l'espace des attributs à différents groupes de décisions partielles obtenus à partir des sorties de la source avec les données générées illustrées sur la Figure 1.

Notons qu'il peut arriver que lors de la phase de test, une décision partielle n'ait pas été rencontrée dans la phase d'apprentissage, c'est-à-dire qu'aucune des instances dans \mathcal{L}_2 n'a conduit à ce que la source fournisse une sortie conduisant à cette décision partielle. Dans cette situation, nous avons choisi la meilleure correction pour l'ensemble des instances dans \mathcal{L}_2 . Dans nos expériences présentées ici, ce cas s'est très rarement produit.

Tableau 3 – Données UCI utilisées.

Bases	# inst.	# attr.	# class.
Breast ca.	569	31	2
Glass	214	10	6
Heberman	306	3	2
Ionosphere	350	34	2
Iris	150	4	3
Liver	345	6	2
Lympho.	146	18	3
Pima	768	8	2
Red wine	1599	12	6
Sonar	208	60	2
Transf.	748	3	2
Vehicles	846	19	4
Vertebral	310	6	3

Les performances selon E_{pl} (6) et les moyennes des mesures d'utilité u_{65} [21] des décisions partielles obtenues à partir des sorties corrigées sont regroupées dans le Tableau 4. La mesure d'utilité u_{65} , introduite par Zaffalon et al. [21], permet d'accorder une plus grande utilité à des décisions partielles imprécises mais correctes de taille n , c'est-à-dire des décisions égales à un ensemble de n singletons dont l'un est la vraie classe, qu'à des décisions précises, en faveur d'un singleton, seulement aléatoirement correctes avec une probabilité de $\frac{1}{n}$. Formellement, la valeur u_{65} d'une décision partielle d , éventuellement en faveur d'un ensemble de singletons, est définie par

$$u_{65}(x) = 1.6x - 0.6x^2 \quad (10)$$

où x est la valeur appelée "fiabilité affaiblie" (*discounted accuracy*) de d , et est définie par $\frac{\mathbb{I}(\omega \in d)}{|d|}$, avec \mathbb{I} la fonction indicatrice, ω la vraie classe de l'instance, et $|d|$ le nombre d'éléments dans d .

Comme on peut le voir dans le Tableau 4, la méthode proposée utilisant les groupes de décisions partielles obtient presque dans chaque situation de meilleurs résultats que l'apprentissage précédent utilisant seulement CD ou CR

ou CN.

La source est censée être une boîte noire, mais nous souhaitons également voir les performances qu'aurait eues cette source si nous étions capables d'améliorer ses performances en utilisant également les données que nous utilisons pour apprendre les corrections. Ces performances sont données dans la colonne $EkNN+$ du Tableau 4. Nous pouvons voir que la nouvelle méthode d'apprentissage réussit même à améliorer ces résultats pour certaines bases.

Comme Ma et Denœux l'ont résumé dans [10], la relation (5) n'est qu'une relation de préférence ou dominance parmi d'autres, aussi de prochains travaux consisteront à tester ces autres relations possibles (t.q. la dominance faible, la maximalité, . . .) pour calculer de nouveaux groupes de décisions partielles possibles.

Remerciements :

Les auteurs tiennent à remercier les deux relecteurs anonymes pour leurs commentaires constructifs et leurs suggestions qui les ont aidés à clarifier certains points de ce papier et qui les aideront dans leurs recherches futures.

Les recherches de Mme Mutmainah sont soutenues par le programme 5000 docteurs étrangers du ministère indonésien des affaires religieuses (bourse française MORA) / Mrs. Mutmainah's research is supported by the overseas 5000 Doctors program of Indonesian Religious Affairs Ministry (MORA French Scholarship).

Références

- [1] C.M. Bishop. Pattern Recognition and Machine Learning. Springer, New York, United States, 2006.
- [2] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Trans. Syst. Man Cybern., 25(5) : 804-813, 1995.
- [3] T. Denœux. Analysis of evidence-theoretic decision rules for pattern classification. Pattern recognit., 30(7) :1095–1107, 1997.
- [4] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artif. Intell., 172 : 234-264, 2008.
- [5] T. Denœux. Evclass : Evidential distance-based classification, [https://cran.r-project.org/web/packages/evclass/index.html], R package version 1.1.1. 2017.
- [6] T. Denœux. Decision-making with belief functions : A review. Int. J. of Approx. Reason., 109 :87–110, 2019.
- [7] D. Dua, C. Graff. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA : Univer-

Tableau 4 – Performances (valeurs moyennes de E_{pl} , les plus basses étant les meilleures, et utilités moyennes de u_{65} , les plus grandes étant les meilleures) obtenues pour EkNN appris avec \mathcal{L}_1 (désigné par EkNN), EkNN appris avec les groupes \mathcal{L}_1 et \mathcal{L}_2 (désigné par EkNN+), EkNN appris avec \mathcal{L}_1 suivi de corrections (CD, CR, CN) apprises avec \mathcal{L}_2 , et EkNN appris avec \mathcal{L}_1 suivi de la proposition d'utiliser des groupes de décisions partielles apprises avec \mathcal{L}_2 . Les écarts types sont indiqués entre parenthèses. Les meilleures performances sont affichées en gras.

Données	EkNN	EkNN+	CD	CR	CN	Proposition
Valeurs moyennes de E_{pl} (les plus basses sont les meilleures)						
Synthétiques	188.27 (26.48)	181.27 (16.23)	187.91 (26.56)	134.57 (9.83)	174.06 (21.38)	106.76 (7.29)
Breast	3.74 (2.50)	3.59 (2.43)	3.65 (2.35)	3.64 (2.34)	3.56 (2.22)	3.56 (2.21)
Glass	8.16 (3.61)	8.97 (3.19)	8.13 (3.57)	5.11 (1.40)	6.93 (2.04)	4.81 (1.49)
Heberman	8.59 (2.40)	8.79 (2.34)	8.57 (2.32)	5.97 (1.84)	6.61 (1.07)	5.87 (1.83)
Ionosphere	2.31 (1.66)	1.64 (1.18)	2.31 (1.62)	2.19 (1.49)	2.25 (1.52)	1.81 (1.49)
Iris	0.56 (0.87)	0.59 (1.02)	0.56 (0.85)	0.55 (0.82)	0.56 (0.84)	0.56 (0.86)
Lymphography	3.07 (1.44)	2.32 (1.27)	3.08 (1.44)	2.74 (1.24)	3.02 (1.18)	2.74 (1.40)
Liver	12.66 (2.08)	12.21 (1.56)	12.66 (2.02)	8.17 (1.08)	8.31 (0.64)	8.01 (1.12)
Pima	20.01 (3.32)	18.12 (3.21)	20.01 (3.32)	15.67 (2.20)	16.76 (1.57)	14.93 (2.40)
Red wine	162.09 (21.75)	140.60 (13.98)	162.09 (21.75)	48.62 (2.96)	116.83 (4.28)	45.92 (3.44)
Sonar	4.04 (1.43)	3.06 (1.33)	4.05 (1.43)	3.50 (1.08)	3.80 (1.05)	3.30 (1.37)
Transfusion	20.54 (4.91)	19.77 (4.37)	19.43 (4.19)	15.21 (3.45)	15.44 (2.21)	13.56 (2.85)
Vehicles	34.26 (5.66)	28.14 (4.74)	34.26 (5.66)	24.07 (2.35)	32.40 (4.20)	20.69 (2.96)
Vertebral	5.08 (2.29)	4.68 (2.05)	5.04 (2.19)	4.64 (1.89)	4.85 (1.89)	4.31 (1.81)
Valeurs moyennes de u_{65} (les plus hautes sont les meilleures)						
Synthétiques	66.00 (3.00)	66.46 (3.00)	65.96 (3.00)	65.59 (3.00)	64.91 (3.00)	68.20 (3.00)
Breast	92.88 (5.00)	93.32 (5.00)	92.91 (5.00)	92.88 (5.00)	92.90 (5.00)	92.94 (5.00)
Glass	62.51 (14.00)	61.68 (13.00)	62.62(14.00)	63.61 (14.00)	54.96 (15.00)	64.46(14.00)
Heberman	74.41 (8.00)	73.93 (8.00)	74.62 (8.00)	74.39 (10.00)	71.78 (8.00)	74.58 (10.00)
Ionosphere	93.64 (5.00)	95.28 (4.00)	93.62 (5.00)	93.65 (5.00)	93.64 (5.00)	93.75 (5.00)
Iris	95.64 (7.00)	95.79 (8.00)	95.64 (7.00)	95.73 (7.00)	95.64 (7.00)	95.92(7.00)
Liver	65.88 (7.00)	66.98 (6.00)	66.05 (6.00)	66.43 (8.00)	64.70 (3.00)	65.59 (8.00)
Lymphography	76.71 (14.00)	82.92 (13.00)	76.67 (14.00)	76.25 (15.00)	75.61 (13.00)	76.43 (16.00)
Pima	72.49 (5.00)	73.05 (6.00)	72.49 (5.00)	72.87 (6.00)	72.24 (4.00)	72.91 (5.00)
Red wine	45.03 (4.00)	52.80 (4.00)	45.02 (4.00)	57.43 (4.00)	25.00 (0.00)	58.45(4.00)
Sonar	78.31 (9.00)	83.86 (8.00)	78.31 (9.00)	78.78 (10.00)	77.88 (8.00)	78.11(11.00)
Transfusion	72.33 (6.00)	73.49 (6.00)	74.39 (6.00)	74.13 (7.00)	71.08 (7.00)	76.17 (7.00)
Vehicles	60.90 (6.00)	64.67 (6.00)	60.90 (6.00)	60.34 (6.00)	56.44 (6.00)	63.65(6.00)
Vertebral	80.26 (9.00)	82.25 (8.00)	80.48 (9.00)	80.25 (9.00)	80.83 (9.00)	81.82 (9.00)

- sity of California, School of Information and Computer Science, 2019.
- [8] D. Dubois, H. Prade. A set-theoretic view of belief functions : logical operations and approximations by fuzzy sets. *Int. J. Gen. Syst.*, 12(3) :193-226, 1986.
- [9] Z. Elouedi, K. Mellouli, P. Smets. Assessing sensor reliability for multisensor data fusion within the transferable belief model. *IEEE Trans. Syst. Man Cybern. B*, 34(1) :782–787, 2004.
- [10] L. Ma, T. Denœux. Partial classification in the belief function framework. *Knowledge-Based Syst.*, 214 : article 106742, 2021.
- [11] D. Mercier, B. Quost, T. Denœux. Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextual Discounting. *Inf. Fusion*, 9(2) :246–258, 2008.
- [12] D. Mercier, G. Cron, T. Denœux, M.-H. Masson. Decision fusion for postal address recognition using belief functions. *Expert Syst. Appl.*, 36(3) :5643–5653, 2009.
- [13] D. Mercier, E. Lefèvre, F. Delmotte. Belief functions contextual discounting and canonical decompositions. *Int. J. Approx. Reason.*, 53(2) :146-158, 2012.
- [14] F. Pichon, D. Dubois, T. Denœux. Relevance and truthfulness in information correction and fusion. *Int. J. Approx. Reason.*, 53(2) :159-175, 2012.
- [15] F. Pichon, D. Mercier, E. Lefèvre, F. Delmotte. Proposition and learning of some belief function contextual correction mechanisms. *Int. J. Approx. Reason.*, 72 :4-42, 2016.
- [16] G. Shafer. A mathematical theory of evidence. Princeton University Press, Princeton, N.J, 1976.

- [17] P. Smets, Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem, *Int. J. Approx. Reason.*, 9(1) : 1-35, 1993.
- [18] P. Smets, R. Kennes. The Transferable Belief Model. *Artif. Intell.*, 66(2) : 191-234, 1994.
- [19] M.C. Troffaes. Decision making under uncertainty using imprecise probabilities. *Int. J. Approx. Reason.*, 45(1) :17–29, 2007.
- [20] G. Yang, S. Destercke, M.H. Masson. Nested Dichotomies with probability sets for multi-class classification. *ECAI*, pages 363–368, 2014.
- [21] M. Zafallon, G. Corani, D.-D. Mauá. Evaluating credal classifiers by utility-discounted predictive accuracy. *Int. J. Approx. Reason.*, 53(8) : 1282-1301, 2012.