

Améliorer un groupe de sources crédibilistes avec des corrections contextuelles

S. Mutmainah¹

S. Hachour²

F. Pichon²

D. Mercier²

¹ UIN Sunan Kalijaga, Yogyakarta, Indonesia.

² Univ. Artois, EA 3926 LGI2A, Béthune, F-62400, France.

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, nous étudions l'intérêt d'apprendre un groupe de sources crédibilistes en utilisant des corrections contextuelles, ce qui revient à optimiser une combinaison conjonctive plutôt qu'optimiser chaque source individuellement. Plusieurs expériences sur des données synthétiques et réelles de l'UCI montrent l'intérêt de l'approche.

Mots-clés :

Fusion d'informations, Fonctions de croyance, Corrections contextuelles, Optimisation.

Abstract:

In this paper, we study the interest of learning a group of evidential sources using contextual corrections, which amounts to optimizing a conjunctive combination rather than optimizing each source individually. Several experiments on synthetic and real UCI data show the interest of the approach.

Keywords:

Information Fusion, Belief functions, Contextual corrections, Optimization.

1 Introduction

La fusion d'informations [1, 10] permet, en combinant différentes sources d'informations hétérogènes, d'obtenir une meilleure compréhension (éventuellement plus complète, plus précise) de la situation en cours d'évaluation.

La théorie de Dempster-Shafer des fonctions de croyance [16, 2, 15] permet de représenter l'imprécision et l'incertitude d'une information, ainsi que la manipulation de ces informations imparfaites. Elle constitue un cadre mathématique riche et flexible, largement utilisé pour modéliser un problème de fusion d'informations [8, 14]. Un schéma classique de

fusion consiste à modéliser les sorties individuelles des sources aussi finement que possible pour en faire des éléments d'information indépendants et fiables afin de pouvoir les combiner en utilisant la règle de combinaison conjonctive (c'est-à-dire la règle de Dempster non normalisée). La fiabilité des sorties des sources peut être assurée en utilisant l'opération d'affaiblissement [16, 13] ou des corrections plus fines telles que les corrections contextuelles [11, 12] parmi lesquels l'affaiblissement contextuelle (noté CD, *Contextual Discounting*), permettant d'affaiblir une information et qui généralise l'affaiblissement, ou le renforcement contextuel (noté CR, *Contextual Reinforcement*), qui peut renforcer la sortie d'une source, ou le reniement contextuel (notée CN, *Contextual Negating*), capable de nier ce qu'une source indique.

Dans l'opération d'affaiblissement [16], la fiabilité de la source, fournissant une fonction de masse m , est prise en compte à l'aide d'un réel $\beta \in [0, 1]$ quantifiant le degré de croyance sur le fait que la source est fiable, et la fonction de masse corrigée est notée ${}^\beta m$. Dans les mécanismes de correction contextuelle (CD, CR et CN), l'imperfection de la source, son biais au sens large, est modélisée à l'aide d'un vecteur $\beta \in [0, 1]^C$, avec $C \leq 2^K$ et K le nombre d'éléments de l'univers (des détails se trouvent dans [12]). La fonction de masse corrigée résultante est également notée ${}^\beta m$ pour plus de simplicité.

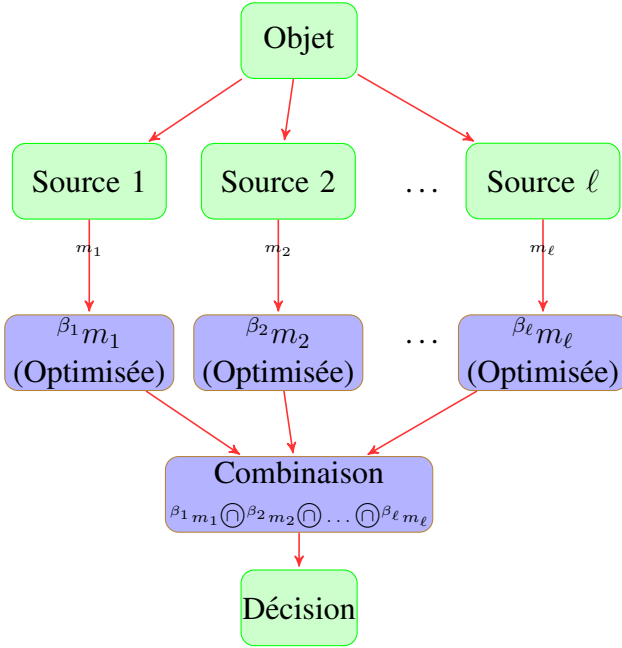


FIGURE 1 – Fusion avec corrections individuelles (Schéma 1).

Si de plus, on dispose d'un ensemble d'apprentissage composé des sorties d'une source, exprimées sous forme de fonctions de masse, concernant les classes de n objets o_i , $i \in \{1, \dots, n\}$ la classe réelle (appartenant à l'univers) de chaque objet étant connue, alors il est possible [9, 11] de trouver les paramètres optimaux β , c'est-à-dire d'apprendre les paramètres β minimisant une mesure de divergence entre les sorties corrigées et les vérités terrain.

Ce schéma classique de fusion d'informations est illustré sur la Figure 1.

Une autre idée, illustrée sur la Figure 2, consiste à apprendre directement une combinaison conjonctive optimisée au lieu d'optimiser chaque source individuellement. Cette idée a été mentionnée dans [9] pour l'opération d'affaiblissement classique et dans [11] pour CD.

Dans ce papier, nous utilisons des classifieurs comme sources d'informations et explorons cette idée d'optimiser directement les performances de la combinaison en utilisant des corrections parmi CD, CR et CN.

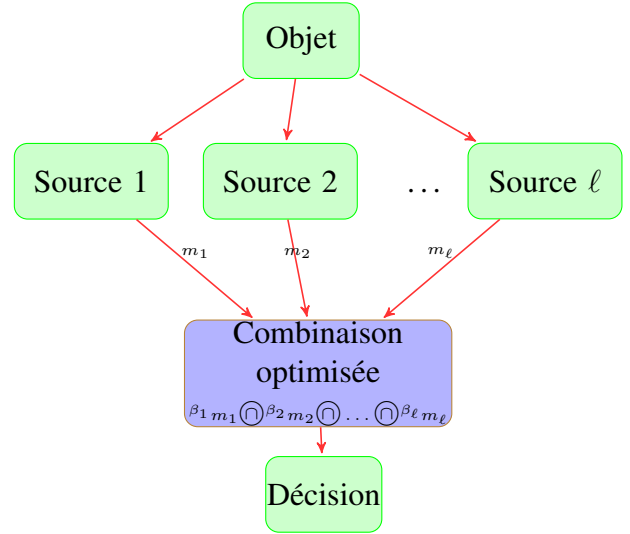


FIGURE 2 – Fusion avec corrections globales (Schéma 2).

Ce papier est organisé de la manière suivante. Les notations et concepts utilisés sont rappelés dans la Section 2. L'apprentissage des corrections contextuelles d'un groupe de classifieurs crédibilistes est présenté dans la Section 3. Des expériences sur des données synthétiques et réelles montrant l'intérêt de l'approche sont exposées dans la Section 4. Enfin, une conclusion est donnée dans la Section 5.

2 Fonctions de croyance : notations et concepts utilisés

2.1 Concepts de base

Les concepts de base sont brièvement rappelés. Des détails sur cette théorie peuvent être trouvés par exemple dans [16, 13, 5].

L'univers Ω est un ensemble fini composé de K éléments $\omega_1, \dots, \omega_K$.

Nous considérons une question Q d'intérêt dont la réponse est dans Ω .

Une information au regard de cette réponse peut être représentée par une fonction de masse (FM)

m définie de 2^Ω dans $[0, 1]$, t.q.

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (1)$$

Le réel $m(A)$ représente la part de croyance allouée au fait que la réponse cherchée est dans A .

Un sous ensemble A de Ω t.q. $m(A) > 0$ est appelé un élément focal (EF) de m .

Une FM catégorique, ou logique, a un seul EF $A \subseteq \Omega$ et est notée m_A . Ainsi nous avons $m_A(A) = 1$. En particulier, m_Ω représente l'ignorance totale.

Une FM m est en correspondance biunivoque avec une fonction de croyance Bel et une fonction de plausibilité Pl , étant respectivement définies pour tout $A \subseteq \Omega$ par :

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

et

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

avec $\bar{A} = \Omega \setminus A$.

La *fonction contour* pl correspond à la restriction de la fonction de plausibilité aux singletons de Ω , elle est définie pour tout $\omega \in \Omega$ par :

$$pl(\omega) = Pl(\{\omega\}). \quad (4)$$

Deux FM fiables et indépendantes m_1 et m_2 définies sur le même univers Ω peuvent être combinées en utilisant la règle de combinaison conjonctive (RCC), ou règle de Dempster non normalisée, définie pour tout $A \subseteq \Omega$ par :

$$\begin{aligned} (m_1 \odot m_2)(A) &= m_1 \odot_2(A) \\ &= \sum_{B \cap C = A} m_1(B) \cdot m_2(C). \end{aligned} \quad (5)$$

2.2 Corrections

Une information, représentée par une FM m , fiable avec un degré $\beta = 1 - \alpha \in [0, 1]$ peut

être affaiblie en utilisant l'opération d'affaiblissement suivante

$${}^\beta m = \beta m + \alpha m_\Omega \quad (6)$$

La fonction de contour associée à cette FM affaiblie ${}^\beta m$ (6) vérifie pour tout $\omega \in \Omega$, ${}^\beta pl(\omega) = 1 - (1 - pl(\omega))\beta$, avec pl la fonction de contour de m (Des détails peuvent être trouvés par exemple dans [13, 11, 12]).

Dans le tableau 1, sont indiquées les fonctions de contour de l'affaiblissement contextuel (CD), du renforcement contextuel (CR) et du reniement contextuel (CN) d'une FM m correspondant à un choix spécifique des paramètres : $C = K$, $\beta_\Omega \in [0, 1]$ pour tout $\omega \in \Omega$, pour chaque correction contextuelle. Les raisons pour lesquelles nous nous limitons à $C = K$ et les définitions de ces paramètres K pour chaque correction contextuelle peuvent être trouvées dans [12, Section 8].

TABLEAU 1 – Fonction de contour de chaque correction contextuelle d'une FM m donnée pour tout $\omega \in \Omega$. Chaque paramètre β_ω peut varier dans $[0, 1]$.

Corr.	Fonctions de contour
CD	${}^\beta pl(\omega) = 1 - (1 - pl(\omega))\beta_\omega$
CR	${}^\beta pl(\omega) = pl(\omega)\beta_\omega$
CN	${}^\beta pl(\omega) = 0.5 + (pl(\omega) - 0.5)(2\beta_\omega - 1)$

Comme nous l'avons rappelé dans l'introduction, si pour une source nous disposons d'un ensemble d'apprentissage contenant ses sorties, c'est-à-dire des fonctions de masse $m\{o_i\}$ exprimant les opinions de la source au regard des classes de n objets o_i , $i \in \{1, \dots, n\}$, la classe réelle de chaque objet étant connue, nous pouvons alors calculer les paramètres β des corrections CD, CR et CN en optimisant la mesure suivante de divergence entre les sorties corrigées (par CD, CR ou CN) de la source et les

classes réelles des objets

$$E_{pl}(\beta) = \sum_{i=1}^n \sum_{k=1}^K (\beta pl\{o_i\}(\{\omega_k\}) - \delta_{i,k})^2, \quad (7)$$

où $\beta pl\{o_i\}$ est la fonction de contour concernant la classe de l'objet o_i corrigée avec un vecteur $\beta = (\beta_\omega \in [0, 1], \omega \in \Omega)$ et $\delta_{i,k}$ est la fonction indicatrice des classes réelles des objets o_i , $i \in \{1, \dots, n\}$, c'est-à-dire $\delta_{i,k} = 1$ si la classe de l'objet o_i est ω_k , sinon $\delta_{i,k} = 0$.

Cette mesure E_{pl} donne, pour chaque correction CD, CR et CN, un problème d'optimisation quadratique sous contraintes linéaires qui peut être résolu efficacement.

3 Apprentissage d'un groupe de sources crédibilistes

Lorsque plusieurs sources sont disponibles, au lieu d'apprendre les meilleurs paramètres de correction individuellement pour chaque source en sachant que les fonctions de masse sorties par ces sources vont être ensuite combinées, il est possible d'optimiser directement la combinaison des fonctions de masse corrigées.

Avec ℓ sources à combiner, nous avons ℓ vecteurs $\beta_1, \dots, \beta_\ell$ - chacun possiblement associé soit à CD, soit à CR, soit à CN - pouvant être obtenus en minimisant la mesure de divergence suivante

$$E_{pl}(\beta_1, \dots, \beta_\ell) = \sum_{i=1}^n \sum_{k=1}^K (\beta_1 pl_1\{o_i\}(\{\omega_k\}) \times \dots \times \beta_\ell pl_\ell\{o_i\}(\{\omega_k\}) - \delta_{i,k})^2 \quad (8)$$

En effet, après la combinaison conjonctive, la plausibilité de chaque singleton est égale au produit des ℓ plausibilités données par les sources à ce singleton.

Optimiser (7) pour chaque classifieur ou (8) n'est pas la même chose car un classifieur peut être utilisé de manière différente s'il est utilisé seul ou à travers un collectif.

Cette approche présente toutefois un inconvénient : l'optimisation de (8) n'est plus un

problème d'optimisation d'une forme quadratique sous contraintes linéaires, elle peut être minimisée à l'aide d'une procédure standard d'optimisation non linéaire sous contraintes atteignant un éventuel minimum local.

Avec trois mécanismes possibles (CD, CR et CN), qui peuvent être appliqués sur chaque source, et ℓ sources, le nombre d'optimisations à entreprendre devient également plus grand. En effet, nous avons pour le premier schéma utilisant des corrections individuelles (cf Figure 1) $3 \times \ell$ corrections possibles à tester, tandis que pour ce deuxième schéma optimisant la combinaison (cf Figure 2), nous avons 3^ℓ corrections possibles à tester.

Par exemple, considérons le cas de deux sources S_1 et S_2 ($\ell = 2$). Pour les optimisations individuelles, nous avons pour chaque source 3 optimisations à entreprendre, en utilisant (7), pour savoir quelle correction entre CD, CR et CN conserver pour chaque source, et donc finalement $3 \times 2 = 6$ optimisations au total de (7). Alors que, pour l'optimisation directe de la combinaison utilisant (8), nous devons comparer toutes les associations possibles de corrections pour les sources S_1 et S_2 (CD-CD, CD-CR, CD-CN, CR-CD, CR-CR, CR-CN, CN-CD, CN-CR et CN-CN) conduisant alors à un cadre plus riche de corrections possibles, mais avec plus de comparaisons à faire, $3^2 = 9$ dans ce scénario.

Dans la section suivante, nous montrons à l'aide de plusieurs expériences sur des données synthétiques et réelles, que ce deuxième schéma peut avoir un intérêt en raison de ses performances.

4 Expérimentations

Pour tester ces schémas (corrections individuelles - Figure 1 - vs correction globale - Figure 2), plusieurs expériences numériques menées sur des ensembles de données synthétiques et réelles en utilisant deux classifieurs crédibilistes sont présentées dans

cette section.

Le premier classifieur est le k plus proche voisins crédibiliste (EkNN) [3, 6] avec $k = 5$. Le deuxième classifieur choisi est le réseau de neurones crédibiliste (ENN) [4, 6] avec un nombre de prototypes $np = 5$. Le choix des valeurs de paramètres $k = 5$ et $np = 5$ donne des classifieurs raisonnables, l'idée n'est pas d'optimiser chaque classifieur sur chaque base de test mais de considérer ces classifieurs comme des boîtes noires et voir comment les utiliser au mieux.

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 ou ENN) ;
- Une validation croisée à 10 blocs est ensuite effectuée sur la seconde moitié des données avec 9 blocs (\mathcal{L}_2) pour apprendre la meilleure correction, et 1 bloc pour le test.

Les données synthétiques, illustrées sur la Figure 3, ont été générées par une distribution normale multivariée composée de 2 caractéristiques, 900 objets et 3 classes avec les moyennes $\mu_1 = (0, 2)$, $\mu_2 = (1, 3)$, $\mu_3 = (2, 2)$ et les matrices de covariance suivantes pour chaque classe :

$$\Sigma_1 = 0,1I \quad (9)$$

$$\Sigma_2 = 0,5I \quad (10)$$

$$\Sigma_3 = \begin{bmatrix} 0,3 & -0,15 \\ -0,15 & 0,3 \end{bmatrix} \quad (11)$$

où I est la matrice identité de taille 2×2 .

Les données réelles utilisées proviennent de la base UCI [7] (Détails dans le tableau 2).

Les résultats sont résumés dans le tableau 3 en utilisant comme mesure de performance E_{pl} (7), c'est-à-dire l'erreur quadratique entre la fonction de contour résultant de la combinaison et

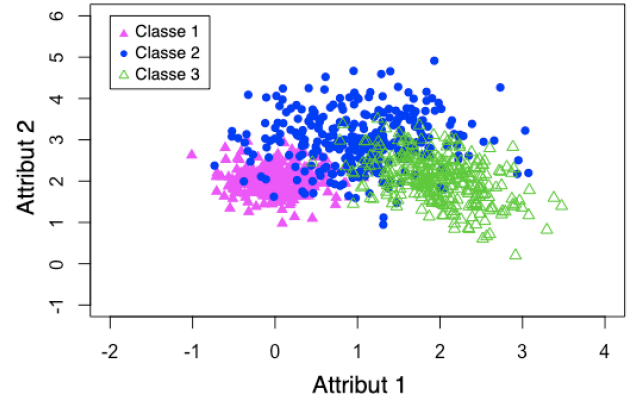


FIGURE 3 – Données générées.

TABEAU 2 – Données UCI utilisées.

Bases	# inst.	# attr.	# class.
Haberman	306	3	2
Iris	150	4	3
Glass	214	10	6
Ionosphere	350	34	2
Lymphography	140	18	3
Liver	345	6	2
Pima	768	8	2
Sonar	208	60	2
Transfusion	748	3	2
Vehicle	846	19	4
Vertebral	310	6	3

la fonction indicatrice des vérités des objets de l'ensemble de test.

On peut voir dans le tableau 3 que le second schéma optimisant la combinaison atteint de meilleures performances selon E_{pl} (7) que le premier schéma combinant les optimisations individuelles.

Nous avons également voulu mettre en évidence l'intérêt éventuel de tirer profit de l'utilisation de plusieurs corrections différentes et c'est pourquoi les performances du schéma 2 avec seulement CD, seulement CR et seulement CN ont également été exposées pour comparaison. L'utilisation d'un seul type de correction peut limiter les performances.

TABLEAU 3 – Performances (valeurs moyennes de $E_{pl}(7)$), les plus faibles étant les meilleures, obtenues à partir de deux sources (EkNN et ENN) avec le schéma 1 (meilleures corrections individuelles), le schéma 2 (meilleure combinaison paramétrée), le schéma 2 avec seulement CD, seulement CR et seulement CN pour souligner l'intérêt d'utiliser éventuellement plusieurs corrections distinctes. Les écarts types sont indiqués entre parenthèses. En gras, la meilleure performance pour chaque ensemble de données.

Données	Schéma 1	Schéma 2	Schéma 2-CD	Schéma 2-CR	Schéma 2-CN
Générées	5.244 (2.415)	4.250 (2.106)	4.251 (2.107)	5.142 (2.555)	5.134 (2.521)
Haberman	6.737 (1.672)	5.514 (1.734)	5.938 (2.224)	5.696 (2.079)	5.561 (1.866)
Iris	0.561 (0.614)	0.470 (0.730)	0.461 (0.700)	0.495 (0.621)	0.495 (0.621)
Glass	6.009 (0.743)	4.634 (1.371)	4.786 (1.656)	4.708 (1.415)	4.719 (1.264)
Ionosphere	2.911 (0.865)	2.075 (0.981)	2.075 (0.981)	2.439 (0.902)	2.439 (0.902)
Lympho	2.743 (0.937)	2.245 (1.044)	2.237 (1.046)	2.317 (1.101)	2.324 (1.071)
Liver	9.513 (0.902)	7.569 (1.187)	7.894 (1.520)	7.927 (1.500)	7.789 (1.327)
Pima	16.384 (1.995)	12.411 (2.437)	12.461 (2.506)	13.135 (2.754)	13.110 (2.702)
Sonar	4.360 (0.834)	3.147 (0.916)	3.142 (0.966)	3.481 (1.041)	3.481 (0.994)
Transfusion	16.494 (2.257)	13.221 (2.242)	15.586 (3.062)	15.192 (3.046)	13.520 (2.178)
Vehicles	30.698 (1.165)	21.523 (2.198)	22.381 (2.925)	22.452 (2.699)	21.590 (2.134)
Vertebral	5.260 (1.656)	4.007 (1.479)	4.010 (1.489)	4.676 (1.759)	4.565 (1.555)

On peut observer qu'il arrive que le schéma 2 avec uniquement CD (Schéma 2-CD) obtienne de légères meilleures performances (Bases Iris, Lympho et Sonar) que le schéma 2 testant toutes les combinaisons dont CD-CD. Ceci peut s'expliquer de plusieurs manières non exclusives :

1. La meilleure configuration sur l'ensemble d'apprentissage n'est pas forcément la meilleure sur l'ensemble de test.
2. L'optimisation sur l'ensemble d'apprentissage n'est que locale.
3. La mesure de performance $E_{pl}(7)$ est un peu favorable à CD (Détails dans [12, Section 8.5.1]).

Comme attendu, l'inconvénient pour atteindre ces performances est un temps plus long pour apprendre les paramètres comme le montre le Tableau 4. Avec seulement deux sources, ce temps reste raisonnable. Si le nombre de sources devait devenir trop important, il serait certainement nécessaire de voir si le schéma 2 est toujours applicable dans un temps raison-

nable.

5 Conclusion

Dans ce papier, nous avons illustré par des expérimentations les intérêts d'utiliser différentes corrections contextuelles pour optimiser la combinaison conjonctive des sorties d'un groupe de sources crédibilistes. Nous avons également donné des éléments sur les limites de cette stratégie lorsque le nombre de sources à combiner augmente, cette limite restant à préciser. Une perspective intéressante et d'actualité sera d'étudier la possibilité d'utiliser ces schémas dans un apprentissage de bout en bout d'un groupe de classificateurs crédibilistes profonds dans la lignée des travaux de Tong et al. par exemple [17].

Remerciements :

Les auteurs tiennent à remercier les relecteurs anonymes pour leurs commentaires constructifs et leurs remarques pertinentes qui les aideront dans leurs recherches futures.

TABLEAU 4 – Moyennes en secondes des temps consommés sur un macbook Air M1 3.2 GHz 8 GB RAM pour la phase d'apprentissage pour le schéma 1 et le schéma 2 lors de ces expérimentations sur chacune des bases. Les écarts types sont indiqués entre parenthèses.

Données	Schéma 1	Schéma 2
Générées	0.0477 (0.0116)	42.7417 (16.4333)
Haberman	0.0120 (0.0079)	11.4312 (1.3702)
Iris	0.0059 (0.0010)	5.9721 (1.1307)
Glass	0.0084 (0.0012)	18.5167 (3.9630)
Ionosphere	0.0120 (0.0020)	7.9370 (0.4878)
Lympho	0.0074 (0.0110)	6.5583 (0.7286)
Liver	0.0125 (0.0027)	13.5109 (1.0809)
Pima	0.0300 (0.0184)	22.2145 (3.2915)
Sonar	0.0070 (0.0013)	4.8880 (0.5885)
Transfusion	0.0281 (0.0078)	27.4938 (7.9787)
Vehicles	0.0596 (0.0223)	1.4006 (0.0770)
Vertebral	0.0124 (0.0017)	15.9031 (2.0436)

Références

- [1] I. Bloch, A. Hunter, A. Appriou, A. Ayoun, S. Benferhat, L. Cholvy, R. Cooke, F. Cuppens, D. Dubois, H. Fargier. Fusion : General concepts and characteristics. *Int. J. Intell. Syst.*, 16(10) :1107–1136, 2001.
- [2] A.P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *Ann. Math. Stat.*, 38 :325–339, 1967.
- [3] T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. Syst. Man Cybern.*, 25(5) :804-813, 1995.
- [4] T. Denœux. A neural network classifier based on Dempster-Shafer theory. *IEEE Trans. Syst. Man Cybern.*, 30(2) :131–150, 2000.
- [5] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artif. Intell.*, 172 :234-264, 2008.
- [6] T. Denœux. Evclass : Evidential distance-based classification, [<https://cran.r-project.org/web/packages/evclass/index.html>], R package version 1.1.1. 2017.
- [7] D. Dua, C. Graff. UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA : University of California, 2019.
- [8] D. Dubois, W. Liu, J. Ma, H. Prade. The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. *Inf. Fusion*, 32(A) :12–39, 2016.
- [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.I. Kuncheva. *Combining Pattern Classifiers : Methods and Algorithms*, 2nd Edition. Wiley, Hoboken, New Jersey, United States, 2014.
- [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] 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.
- [13] P. Smets. Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem, *Int. J. Approx. Reason.*, 9(1) : 1-35, 1993.
- [14] P. Smets. Analyzing the combination of conflicting belief functions, *Inf. Fusion*, 8(4) :387–412, 2007.
- [15] P. Smets, R. Kennes. The Transferable Belief Model. *Artificial Intelligence.*, 66(2) :191-234, 1994.
- [16] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J, 1976.
- [17] Z. Tong, P. Xu, T. Denœux. Fusion of Evidential CNN Classifiers for Image Classification. In T. Denœux, E. Lefèvre, Z. Liu and F. Pichon (Eds), *Belief Functions : Theory and Applications*, Springer International Publishing, pages 168-176, 2021.