Apprentissage de corrections contextuelles crédibilistes à partir de données partiellement étiquetées en utilisant la fonction de contour

Siti Mutmainah, Frédéric Pichon, David Mercier

Apprentissage de corrections contextuelles crédibilistes à partir de données partiellement étiquetées en utilisant la fonction de contour

Learning evidential contextual corrections from soft labelled data using the contour function

S. Mutmainah[1,2] F. Pichon[1] D. Mercier[1]
[1] Univ. Artois, EA 3926 LGI2A, Béthune, F-62400, France.
[2] UIN Sunan Kalijaga, Yogyakarta, Indonesia.

siti.mutmainah@uin-suka.ac.id
frederic.pichon@univ-artois.fr
david.mercier@univ-artois.fr

Résumé :

Dans ce papier, nous proposons d'apprendre les paramètres de mécanismes de corrections contextuelles crédibilistes à partir de données d'apprentissage partiellement étiquetées, c'est-à-dire de données où la vraie classe de chaque objet n'est connue que partiellement, en optimisant une mesure de différence entre les valeurs de la fonction de contour corrigée et la vérité terrain également représentée par une fonction de contour. Les avantages de cette méthode sont illustrés par des tests sur des données synthétiques et réelles.

Mots-clés :

Fonctions de croyance, Corrections contextuelles, Apprentissage, Étiquettes partielles.

Abstract:

In this paper, we propose to learn the parameters of evidential contextual correction mechanisms from a learning set composed of partially labelled data (soft labels), i.e. data where the true class of each object is only partially known, by optimizing a measure of discrepancy between the values of the corrected contour function and the ground truth also represented by a contour function. The advantages of this method are illustrated by tests on synthetic and real data.

Keywords:

Belief functions, Contextual corrections, Learning, Soft labels.

1 Introduction

Dans la théorie des fonctions de croyance [13, 15], la correction d'une source d'information, un capteur par exemple, peut être réalisée classiquement par l'opération d'affaiblissement (*discounting*) introduite par Shafer [13], mais aussi par des méthodes dites contextuelles [8, 11] prenant en compte plus finement la qualité d'une source. Nommément il s'agit dans ce papier des mécanismes d'affaiblissement, renforcement et reniement contextuels [11].

Ces mécanismes peuvent se dériver des notions de fiabilité (ou pertinence) qui concerne la compétence d'une source vis-à-vis de la question d'intérêt, et de sincérité [10, 11] qui indique la capacité de la source à dire ce qu'elle sait, on peut également parler de biais de la source. L'affaiblissement contextuel (*contextual discounting* (CD)) est une extension de l'affaiblissement qui correspond à une source partiellement fiable et totalement sincère. Le reniement contextuel (*contextual negating* (CN)) est une extension du reniement qui correspond au cas d'une source totalement pertinente mais partiellement sincère, le cas extrême étant la négation d'une information. Enfin le renforcement contextuel (*contextual reinforcement* (CR)) est une extension du renforcement une opération duale de l'affaiblissement [9, 11].

Dans ce papier, nous abordons le problème de l'apprentissage des paramètres de ces mécanismes de correction à partir de données partiellement étiquetées, c'est-à-dire de données où la vraie classe de chaque objet n'est connue que partiellement, cette connaissance étant modélisée en utilisant une fonction de contour dans notre cas. Une méthode pour

apprendre ces corrections à partir de données étiquetées, où la vérité est connue sûrement et précisément pour chaque élément de l'ensemble d'apprentissage, a déjà été introduite dans [11] ; elle consiste à minimiser une mesure de différence (ou divergence) entre les valeurs de la fonction de contour corrigée et la vérité. Nous montrons ici que cette même mesure peut être reprise pour apprendre à partir de données partiellement étiquetées, et illustrons avec des tests sur des données synthétiques et réelles l'avantage de cette méthode pour 1) améliorer un classifieur même si les données sont seulement partiellement étiquetées et 2) obtenir de meilleures performances en apprenant directement à partir de ces vérités partielles (*soft labels*) plutôt qu'à partir de données de vérités dures (*hard labels*) approchant ces vérités partielles (*soft labels*) uniquement disponibles.

Ce papier est organisé de la manière suivante. Dans la Section 2, les concepts de base utilisés et les notations employées sont présentés. Puis, dans la Section 3, sont abordés les trois corrections contextuelles utilisées (affaiblissement, renforcement et reniement contextuels) ainsi que leurs apprentissages à partir de données étiquetées et la question de comment nous proposons de reprendre cet apprentissage avec des données partiellement étiquetées. Les tests de cette méthode sont alors exposés dans la Section 4. Enfin, une discussion des résultats et des pistes de recherches futures sont présentées dans la Section 5.

2 Fonctions de croyance : Concepts de base utilisés et notations

Seuls les concepts de base utilisés et les notations employées sont présentés dans cette section (voir par exemple [13, 15, 3] pour plus de détails sur le cadre des fonctions de croyance).

À partir d'un cadre de discernement $\Omega = \{\omega_1, ..., \omega_K\}$, une fonction de masse (FM), notée m^Ω ou m en l'absence d'ambiguïté, est définie de 2^Ω dans $[0, 1]$, et vérifie $\sum_{A \subseteq \Omega} m^\Omega(A) = 1$.

Les éléments focaux d'une FM m sont les sous ensembles A de Ω tels que $m(A) > 0$.

Une FM m est en bijection avec une fonction de plausibilité Pl définie pour tout $A \subseteq \Omega$ par

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \quad (1)$$

La fonction de contour d'une FM m est définie pour tout $\omega \in \Omega$ par

$$\begin{array}{rccc} pl: & \Omega & \rightarrow & [0,1] \\ & \omega & \mapsto & pl(\omega) = Pl(\{\omega\}) \end{array} \quad (2)$$

Elle coïncide avec la plausibilité de m sur tous les singletons de Ω.

La connaissance de la fiabilité d'une source est classiquement prise en compte par l'opération d'affaiblissement (*discounting*) [13, 14]. Supposons une source S fournissant une information représentée par une FM m_S. Avec $\beta \in [0, 1]$ le degré de croyance concernant la fiabilité de la source, l'affaiblissement de m_S est définie par la FM m telle que :

$$m(A) = \beta\, m_S(A) + (1 - \beta) m_\Omega(A)\ , \quad (3)$$

pour tout $A \subseteq \Omega$, et où m_Ω représente l'ignorance totale, c'est-à-dire la FM définie par $m_\Omega(\Omega) = 1$.

Plusieurs justifications de ce mécanisme se trouvent dans [14, 8, 11].

La fonction de contour associée à l'affaiblissement (voir par exemple [11, Prop. 11]) est définie pour tout $\omega \in \Omega$ par :

$$pl(\omega) = 1 - (1 - pl_S(\omega))\beta\ , \quad (4)$$

où pl_S est la fonction de contour issue de la FM m_S fournie par la source.

3 Corrections contextuelles et apprentissage

3.1 Corrections contextuelles d'une fonction de croyance

Pour plus de simplicité, nous rappelons uniquement les expressions des fonctions de contour

résultant de l'application des mécanismes d'affaiblissement, renforcement et reniement contextuels dans le cas de K contextes où K est le cardinal de Ω. Dans [11], il est montré que ces expressions sont suffisamment riches pour minimiser la mesure de différence permettant d'apprendre les paramètres de ces corrections et qui est présentée dans la Section 3.2.

On suppose qu'une source S fournit une information représentée par une FM m_S.

Pour l'affaiblissement contextuel (CD), la fonction de contour résultante de l'affaiblissement de m_S à partir d'un ensemble de contextes composé des singletons de Ω est donné par

$$pl(\omega) = 1 - (1 - pl_S(\omega))\beta_{\{\omega\}} \,, \qquad (5)$$

pour tout $\omega \in \Omega$, avec les K paramètres $\beta_{\{\omega\}}$ pouvant varier dans $[0, 1]$.

Pour le renforcement contextuel (CR) et le reniement contextuel (CN), les fonctions de contour sont données, à partir d'un ensemble de contextes composé des complémentaires des singletons de Ω, respectivement par

$$pl(\omega) = pl_S(\omega)\beta_{\overline{\{\omega\}}} \,, \qquad (6)$$

et

$$pl(\omega) = 0.5 + (pl_S(\omega) - 0.5)(2\beta_{\overline{\{\omega\}}} - 1) \,, \quad (7)$$

pour tout $\omega \in \Omega$, avec les K paramètres $\beta_{\overline{\{\omega\}}}$ pouvant varier dans $[0, 1]$.

3.2 Apprentissage à partir de données étiquetées

Supposons que l'on dispose d'une source d'information fournissant une FM m_S au regard de la vraie classe d'un objet parmi un ensemble Ω de classes possibles.

Si l'on dispose d'un ensemble d'apprentissage composé de n instances (ou objets) dont la vérité (la classe réelle) est connue, on peut apprendre les paramètres d'une correction minimisant une mesure de différence entre la sortie corrigée du classifieur (une correction est appliquée à m_S) et la vérité [6, 8, 11].

La mesure E_{pl} suivante, introduite dans [8], permet d'aboutir à un problème d'optimisation simple (minimisation d'une forme quadratique) pour optimiser les vecteurs $\boldsymbol{\beta}_{CD}$, $\boldsymbol{\beta}_{CR}$ et $\boldsymbol{\beta}_{CN}$ de K paramètres des corrections CD, CR et CN :

$$E_{pl}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}(pl_i(\omega_k) - \delta_{i,k})^2 \,, \qquad (8)$$

où pl_i est la fonction de contour quant à la classe de l'instance i résultant d'une correction contextuelle (CD, CR ou CN) de la FM fournie par la source pour cette instance, et $\delta_{i,k}$ est la fonction indicatrice de la vérité de l'instance i, c'est-à-dire $\delta_{i,k} = 1$ si la classe de l'instance i est ω_k, sinon $\delta_{i,k} = 0$.

3.3 Apprentissage à partir de données partiellement étiquetées

Dans ce papier, nous nous intéressons au cas où la vérité n'est plus donnée précisément par les $\delta_{i,k}$, mais seulement de manière imprécise par une fonction de contour $\tilde{\delta}_i$.

Nous proposons de reprendre l'équation (8) avec cette fois

$$\begin{array}{rccc} \tilde{\delta}_i : & \Omega & \rightarrow & [0,1] \\ & \omega_k & \mapsto & \tilde{\delta}_i(\omega_k) = \tilde{\delta}_{i,k} \end{array} \qquad (9)$$

une fonction de contour informant de la vraie classe ω_k dans Ω de l'instance i.

L'optimisation de la mesure $\tilde{E}_{pl}$ définie par

$$\tilde{E}_{pl}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{k=1}^{K}(pl_i(\omega_k) - \tilde{\delta}_{i,k})^2 \,, \qquad (10)$$

conduit également, pour chaque correction (CD, CR et CN), à un problème d'optimisation des moindres carrés pour lequel il existe des méthodes de résolution efficaces (voir [11, Prop. 12, 14 et 16]).

Par exemple, pour la correction CD, $\tilde{E}_{pl}$ peut s'écrire :

$$\tilde{E}_{pl}(\boldsymbol{\beta}) = \|\boldsymbol{Q\beta} - \tilde{\boldsymbol{d}}\|^2 \quad (11)$$

avec

$$\boldsymbol{Q} = \begin{bmatrix} diag(\boldsymbol{pl_1} - 1) \\ \vdots \\ diag(\boldsymbol{pl_n} - 1) \end{bmatrix}, \tilde{\boldsymbol{d}} = \begin{bmatrix} \tilde{\boldsymbol{\delta_1}} - 1 \\ \vdots \\ \tilde{\boldsymbol{\delta_n}} - 1 \end{bmatrix} \quad (12)$$

où $diag(\boldsymbol{v})$ est une matrice diagonale carrée dont la diagonale est composée des éléments du vecteur $\boldsymbol{v}$, et où pour tout $i \in \{1, \ldots, n\}$, $\tilde{\boldsymbol{\delta}}_{\boldsymbol{i}}$ est le vecteur colonne des valeurs de la fonction de contour $\tilde{\delta}_i$, c'est-à-dire $\tilde{\boldsymbol{\delta}}_{\boldsymbol{i}} = (\tilde{\delta}_{i,1}, \ldots, \tilde{\delta}_{i,K})^T$.

Dans la suite, nous testons cette proposition avec des données générées et réelles.

4 Tests avec des données générées et réelles

Nous exposons premièrement comment nous avons construit nos données avec des étiquettes partielles avant d'aborder les tests.

4.1 Génération de vérités partielles

Il n'est pas aisé de trouver dans la littérature des données partiellement étiquetées, aussi, comme dans [1, 12, 7], nous avons construit nos ensembles de données partiellement étiquetées (soft labels) à partir de vérités parfaites (hard labels) en employant la procédure décrite dans l'Algorithme 1 (où $Bêta, \mathcal{B}$, et $\mathcal{U}$ désignent respectivement des lois Bêta, de Bernoulli et uniforme).

Algorithme 1 Génération des étiquettes partielles (soft labels)

Entrée : étiquettes (hard labels) δ_i où l'entier $k \in \{1, \ldots, K\}$ t.q. $\delta_{i,k} = 1$ est noté k_i
Sortie : étiquettes partielles (soft labels) $\tilde{\delta}_i$

procedure HARDTOSOFTLABELS
  pour chaque instance i **faire**
    Tirer $p_i \sim Bêta(\mu = .5, v = .04)$
    Tirer $b_i \sim \mathcal{B}(p_i)$
    si $b_i = 1$ **alors**
      Tirer $k_i \sim \mathcal{U}_{\{1,\ldots,K\}}$
    $\tilde{\delta}_{i,k_i} \leftarrow 1$
    $\tilde{\delta}_{i,k} \leftarrow p_i$ pour tout $k \neq k_i$

L'Algorithme 1 permet d'obtenir des étiquettes partielles d'autant plus imprécises que la classe la plus plausible est fausse.

4.2 Tests réalisés

Le classifieur crédibiliste utilisé comme source d'information est le classifieur évidentiel des k plus proches voisins (EkNN) introduit par Denœux dans [2]. Il peut être vu comme une boîte noire, nous avons choisi $k = 3$.

Les ensembles de tests considérés sont d'abord un ensemble de données synthétiques à 3 classes généré à partir de 3 distributions normales bivariées d'espérances respectives $\mu_{\omega_1} = (1, 2)$, $\mu_{\omega_2} = (2, 1)$ et $\mu_{\omega_3} = (0, 0)$, avec une matrice de covariance commune Σ telle que :

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}. \quad (13)$$

Pour chaque classe, 100 instances ont été générées, le résultat est illustré sur la Figure 1.

Nous avons ensuite considéré plusieurs jeux de données de la base UCI [5] qui comportent des attributs numériques comme nous utilisons le classifieur EkNN. Ces données réelles sont décrites dans le Tableau 1.

Pour chaque jeu de données, les tests menés ont alors consisté en 10 répétitions d'une séparation

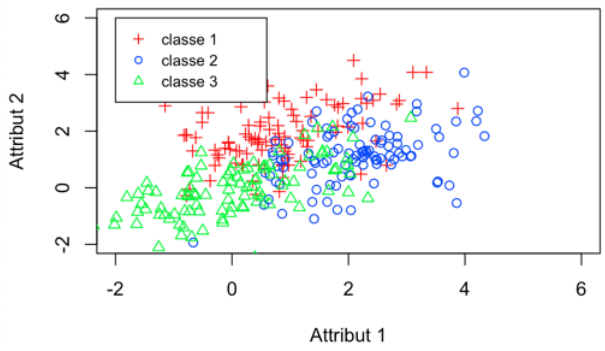

Figure 1 – Illustration du jeu de données générées.

Tableau 1 – Caractéristiques des données réelles UCI utilisées (nombre d'instances sans donnée manquante, nombre de classes, nombre d'attributs numériques utilisés)

Données	# Instances	# Classes	# Attributs
Ionosphere	350	2	34
Iris	150	3	4
Sonar	208	2	60
Vowel	990	11	9
Wine	178	3	13

des données en 10 groupes (10-repeated 10-fold cross validation) où pour chaque séparation :

- le groupe comportant un dixième des données est considéré comme l'ensemble de test (les étiquettes des instances étant rendues imprécises par l'Algorithme 1),
- les 9 autres groupes forment l'ensemble d'apprentissage qui est découpé aléatoirement, à parts égales, en deux groupes :
  - un groupe pour apprendre le classifieur EkNN (appris à partir des vérités certaines),
  - un groupe pour apprendre les paramètres des mécanismes de corrections à partir d'étiquettes partielles (les étiquettes du jeu de données sont rendues imprécises par l'application de l'Algorithme 1).

Pour l'apprentissage des paramètres des corrections contextuelles, deux stratégies sont comparées :

1. Dans une première stratégie, on se propose d'utiliser l'optimisation de l'équation (8) à partir des vérités certaines les plus proches des vérités partielles (la classe la plus plausible est choisie), les corrections associées pour cette apprentissage seront notées CD, CR et CN.
2. Dans la deuxième stratégie, on optimise directement l'équation (10) à partir des étiquettes partielles (cf Section 3.3). Les corrections résultantes de cette apprentissage seront alors notées CDsl, CRsl et CNsl.

Les performances des systèmes (classifieur seul et corrections de ce classifieur suivant les deux stratégies exposées ci-dessus) sont mesurées au sens de $\tilde{E}_{pl}$ (10) où $\tilde{\delta}$ représente la vérité partiellement connue. Cette mesure représente une somme sur les instances de tests des écarts à la vérité recherchée au sens des moindres carrés. Elle atteste du degré de divergence entre la sortie d'un système représentée par sa fonction de

contour et la vérité partiellement connue qui est représentée également par une fonction de contour.

Les performances $\tilde{E}_{pl}$ (10) obtenues sur les données générées et UCI pour le classifieur et ses différentes corrections sont résumées dans les Tableaux 2, 3 et 4 pour chaque type de correction. Les écarts types sont indiqués entre parenthèses.

Tableau 2 – Performances $\tilde{E}_{pl}$ obtenues pour le classifieur seul et le classifieur corrigé par l'affaiblissement contextuel (contextual discounting (CD)) via les deux stratégies. Entre parenthèses les écarts types. "Géné." signifie générées. "Ionos." indique Ionosphère.

Données	EkNN	CD	CDsl
Géné.	23.8 (3.8)	16.6 (2.8)	7.9 (1.5)
Ionos.	16.2 (2.5)	9.6 (2.2)	5.3 (1.0)
Iris	12.5 (2.4)	8.4 (2.1)	3.3 (0.9)
Sonar	7.8 (2.0)	6.3 (1.9)	3.5 (0.9)
Vowel	279 (24)	278 (23)	62 (5)
Wine	13.3 (2.6)	10.4 (2.3)	4.3 (1.0)

Tableau 3 – Performances $\tilde{E}_{pl}$ obtenues pour le classifieur seul et le classifieur corrigé par le renforcement contextuel (contextual reinforcement (CR)) via les deux stratégies. Entre parenthèses les écarts types. "Géné." signifie générées. "Ionos." indique Ionosphère.

Données	EkNN	CR	CRsl
Géné.	23.8 (3.8)	26.8 (3.0)	23.5 (3.7)
Ionos.	16.2 (2.5)	17.2 (1.9)	15.9 (2.3)
Iris	12.5 (2.4)	13.1 (2.0)	12.3 (2.2)
Sonar	7.8 (2.0)	9.0 (1.6)	7.7 (1.9)
Vowel	279 (24)	310 (21)	279 (24)
Wine	13.3 (2.6)	15.0 (2.1)	13.3 (2.5)

On peut remarquer que pour l'affaiblissement contextuel (cf Tableau 2) la deuxième stratégie (CDsl) visant à apprendre directement à partir des étiquettes partielles, permet d'obtenir des écarts à la vérité recherchée plus faibles que la première stratégie (CD) où les paramètres de l'affaiblissement contextuel sont appris à partir d'une vérité approchée. On remarque également que cette stratégie fait mieux que le classifieur sans correction alors que seules des étiquettes partielles sont disponibles.

Tableau 4 – Performances $\tilde{E}_{pl}$ obtenues pour le classifieur seul et le classifieur corrigé par le reniement contextuel (contextual negating (CN)) via les deux stratégies. Entre parenthèses les écarts types. "Géné." signifie générées. "Ionos." indique Ionosphère.

Données	EkNN	CN	CNsl
Géné.	23.8 (3.8)	11.5 (1.6)	9.8 (0.6)
Ionos.	16.2 (2.5)	9.3 (1.3)	8.4 (0.9)
Iris	12.5 (2.4)	6.7 (1.5)	4.8 (0.5)
Sonar	7.8 (2.0)	5.1 (0.8)	5.0 (0.9)
Vowel	279 (24)	240 (21)	65 (5)
Wine	13.3 (2.6)	7.2 (1.6)	5.7 (0.6)

Pour le reniement contextuel (cf Tableau 4), nous avons les mêmes conclusions.

Pour le renforcement contextuel (cf Tableau 3), on a des écarts plutôt plus faibles pour le classifieur mais on a toujours une amélioration de la deuxième stratégie (CRsl) par rapport à la première (CR).

Les performances du renforcement plus faibles que celles des autres mécanismes CD et CN peuvent s'expliquer par le fait que cette correction ne peut que baisser les plaubilités fournies par la source vers zéro [11, Remarque 15], alors que CD peut les augmenter autant que nécessaire jusqu'à un [11, Remarque 14], et CN peut aussi les augmenter [11, Remarque 16]. Dans ces expérimentations, le pouvoir d'affaiblir les informations fournies par la source (i.e. augmenter des plausibilités de singletons) serait avantageux au sens de la mesure de perfor-

mance $\tilde{E}_{pl}$.

5 Discussion et Travaux futurs

Nous avons pu améliorer les performances en utilisant une mesure $\tilde{E}_{pl}$ employant les valeurs des plausibilités retournées par les systèmes pour chaque classe et chaque instance. Avec un simple critère d'erreurs 0-1, où pour chaque instance, nous regardons uniquement la classe la plus plausible et la comparons à la classe vérité, les performances restent les mêmes, la classe la plus plausible restant souvent la même pour chaque classifieur et chaque correction dans le même contexte d'expériences réalisées dans la Section 4.2 (mêmes réglages du classifieur et de l'Algorithme 1 pour générer des vérités partielles notamment).

Pour les travaux futurs, nous souhaitons étudier plus finement les performances des mécanismes en fonction de l'imprécision des étiquettes partielles (i.e. en fonction de la valeur de la probabilité p_i dans l'Algorithme 1).

Nous envisageons également l'utilisation d'autres mesures de performance, qui prendrait aussi en compte complètement la sortie incertaine et imprécise du classifieur. Nous souhaitons par exemple étudier celles introduites par Zaffalon et al. [16].

Il serait aussi possible de tester d'autres classifieurs que l'EkNN. Nous pourrions aussi tester l'avantage de ces mécanismes de corrections dans un contexte de fusion de classifieurs.

Au niveau d'autres méthodes d'apprentissage à partir de données partiellement étiquetées, on peut penser aussi à la vraisemblance crédibiliste introduite par Denœux [4] et déjà mise en place pour développer un EkNN utilisant la correction CD [7].

Remerciements :

Les auteurs remercient les relecteurs anonymes pour toutes leurs remarques pertinentes en particulier pour la suite de ses travaux.

Mrs. Mutmainah's research is supported by the overseas 5000 Doctors program of Indonesian Religious Affairs Ministry (MORA French Scholarship)

Références

[1] E. Côme, L. Oukhellou, T. Denœux, P. Aknin. Learning from partially supervised data using mixture models and belief functions. Pattern Recognition, 42(3) : 334–348, 2009.

[2] T. Denoeux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man, and Cybernetics, 25(5) : 804-813, 1995.

[3] T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial Intelligence, 172 : 234-264, 2008.

[4] T. Denœux. Maximum likelihood estimation from uncertain data in the belief function framework. IEEE Transactions on Knowledge and Data Engineering, 25(1) : 119–130, 2013.

[5] D. Dua, C. Graff. UCI Machine Learning Repository [http ://archive.ics.uci.edu/ml]. Irvine, CA : University of California, School of Information and Computer Science, 2019.

[6] Z. Elouedi, K. Mellouli, P. Smets. The Evaluation of Sensors Reliability and Their Tuning for Multisensor Data Fusion within the Transferable Belief Model. Actes de la 6e conférence European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'2001, pp. 350-361, Toulouse, 2001.

[7] O. Kanjanatarakul, S. Kuson, T. Denœux. An Evidential K-Nearest Neighbor Classifier Based on Contextual Discounting and Likelihood Maximization. Actes de la 5e International Conference on Belief Functions, BELIEF'2018, pp. 155-162, Compiègne, 17-21 septembre, 2018.

[8] D. Mercier, B. Quost, T. Denœux. Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextual Discounting. Information Fusion, 9(2) : 246-258, 2008.

[9] D. Mercier, E. Lefèvre, F. Delmotte. Belief functions contextual discounting and canonical decompositions. International Journal of Approximate Reasoning, 53(2) : 146-158, 2012.

[10] F. Pichon, D. Dubois, T. Denoeux. Relevance and truthfulness in information correction and fusion. International Journal of Approximate Reasoning, 53(2) : 159-175, 2012.

[11] F. Pichon, D. Mercier, E. Lefèvre, F. Delmotte. Proposition and learning of some belief function contextual correction mechanisms. International Journal of Approximate Reasoning, 72 : 4-42, 2016.

[12] B. Quost, T. Denoeux, S. Li. Parametric classification with soft labels using the evidential EM algorithm : linear discriminant analysis versus logistic regression. Adv. Data Analysis and Classification, 11(4) : 659-690, 2017.

[13] G. Shafer. A mathematical theory of evidence. Princeton University Press, Princeton, N.J, 1976.

[14] P. Smets, Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning, 9(1) : 1-35, 1993.

[15] P. Smets, R. Kennes. The Transferable Belief Model. Artificial Intelligence, 66(2) : 191-234, 1994.

[16] M. Zafallon, G. Corani, D.-D. Mauá. Evaluating credal classifiers by utility-discounted predictive accuracy. International Journal of Approximate Reasoning, 53(8) : 1282-1301, 2012.