

# Vers l'utilisation de l'affaiblissement contextuel dans un réseau de neurones crédibiliste

## Towards the use of the contextual discounting in an evidential neural network

S.M. Diène<sup>1</sup>

S. Ramel<sup>1</sup>

F. Pichon<sup>1</sup>

D. Mercier<sup>1</sup>

<sup>1</sup> Univ. Artois, EA 3926 LGI2A, Béthune, F-62400, France.

smamadou.diene@univ-artois.fr

sebastien.ramel@univ-artois.fr

frederic.pichon@univ-artois.fr

david.mercier@univ-artois.fr

### Résumé :

Dans cet article, nous explorons l'idée d'employer l'affaiblissement contextuel, plutôt que l'affaiblissement classique, dans le réseau de neurones crédibiliste introduit par Denœux. L'affaiblissement contextuel étant un raffinement de l'affaiblissement classique, nous montrons comment il serait possible de l'utiliser et comment ses paramètres supplémentaires peuvent être appris pour conduire à des premiers résultats intéressants sur des données générées.

### Mots-clés :

Théorie de Dempster-Shafer, Réseau de neurones crédibiliste, Affaiblissement contextuel

### Abstract:

In this article, we explore the idea of using the contextual discounting, rather than the classical discounting, in the evidential neural network introduced by Denœux. As the contextual discounting is a refinement of the classical discounting, we show how it could be used and how its additional parameters can be learnt to lead to interesting first results on generated data.

### Keywords:

Dempster-Shafer Theory, Evidential Neural Network, Contextual Discounting

## 1 Introduction

L'algorithme des  $K$  plus proches voisins crédibiliste, en anglais *Evidential  $K$ -Nearest Neighbors* (EKNN), introduit par Denœux [2] est un algorithme de classification établi dans le cadre de la théorie de Dempster-Shafer [1, 10], dans lequel chaque voisin d'une instance  $x$  à classer, induit une fonction de masse quant à la vraie classe (inconnue) de l'instance  $x$ . Cette fonction de masse est issue de l'affaiblissement

de l'information relative à la classe du voisin, selon un degré de fiabilité qui est fonction de la distance entre ce voisin et l'instance  $x$ . Les fonctions de masses issues des  $K$  plus proches voisins sont ensuite combinées selon la règle de Dempster, produisant en sortie une fonction de masse quant à la classe de l'instance  $x$ .

Une version adaptative de cet algorithme a ensuite été introduite par Denœux [3] sous la forme d'un réseau de neurones crédibiliste nommé en anglais *Evidential Neural Network* (ENN). Dans cette approche, la fonction de masse concernant la classe de l'instance  $x$  est issue d'un nombre limité de prototypes permettant une classification plus rapide et une réduction significative des besoins en mémoire. Chaque prototype génère une fonction de masse calculée en fonction de sa distance par rapport à l'instance  $x$  à classer, et de son degré d'appartenance à chaque classe. Cet algorithme est implémenté sous forme d'un réseau de neurones avec une couche d'entrée, deux couches cachées et une couche de sortie. Un avantage supplémentaire important d'ENN est qu'il peut être employé en lieu et place de la couche *softmax* d'un réseau de neurones profond afin de rendre ce dernier crédibiliste et ainsi obtenir une quantification plus fine de l'incertitude prédictive [12].

Récemment, Dencoux et al. ont proposé CD-EKNN, pour *Contextual Discounting EKNN* [5], une extension de EKNN employant l'affaiblissement contextuel, en anglais *Contextual Discounting* (CD) [7, 9], à la place de l'affaiblissement classique. L'affaiblissement contextuel est un raffinement de l'affaiblissement classique permettant d'associer un degré de fiabilité spécifique à chaque classe. De plus, CD-EKNN est entraîné en minimisant l'entropie croisée de la fonction de contour normalisée, ce qui permet de réduire la complexité.

L'idée de cet article est d'explorer de tels changements dans ENN en l'enrichissant avec CD afin d'offrir une version adaptative de CD-EKNN, tout en l'entraînant avec l'entropie croisée de la fonction de contour normalisée afin de maîtriser la complexité.

Cet article est organisé de la manière suivante. La Section 2 présente les notions fondamentales de la théorie des fonctions de croyance utilisées dans cet article. La Section 3 introduit CD-ENN après avoir rappelé ENN. La Section 4 montre deux jeux de données générées où CD-ENN améliore ENN. Enfin, la Section 5 donne une conclusion et propose des perspectives.

## 2 Fonctions de croyance et corrections

Les principaux concepts de la théorie des fonctions de croyance utilisés dans cet article sont présentés ici. Des détails peuvent être trouvés par exemple dans [10, 11, 4, 5].

Lorsqu'une question  $Q$  admet une réponse appartenant à un ensemble fini  $\Omega = \{\omega_1, \dots, \omega_c\}$ , l'incertitude sur cette réponse peut être modélisée par une fonction de masse (FM)  $m$  définie sur l'ensemble des parties  $2^\Omega$  de  $\Omega$  dans  $[0, 1]$ , t.q.  $\sum_{A \subseteq \Omega} m(A) = 1$ . Le réel  $m(A)$  peut être interprété comme la part de croyance que la réponse recherchée se trouve dans  $A$ . Un élément focal  $A \subseteq \Omega$  d'une FM  $m$  vérifie  $m(A) > 0$ . À toute FM correspond une

fonction de plausibilité  $Pl$  définies pour tout  $A \subseteq \Omega$  par  $Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$ . La fonction de contour  $pl$  correspond à la restriction de la fonction de plausibilité aux singletons de  $\Omega$ . Autrement dit, pour tout  $\omega \in \Omega$ , on a :  $pl(\omega) = Pl(\{\omega\})$ .

Enfin, il est possible d'approcher une FM par une mesure de probabilité à l'aide de la transformation consistant à normaliser la fonction de contour :

$$plp(\omega) = \frac{pl(\omega)}{\sum_{\omega' \in \Omega} pl(\omega')}, \quad \forall \omega \in \Omega \quad (1)$$

Lorsque deux FM fiables et indépendantes  $m_1$  et  $m_2$  sont définies sur le même ensemble  $\Omega$ , elles peuvent être combinées en utilisant la règle de combinaison de Dempster définie pour tout  $\emptyset \neq A \subseteq \Omega$  par :  $(m_1 \oplus m_2)(A) = m_{1 \oplus 2}(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) \cdot m_2(C)$ , où le terme de conflit  $\kappa$  est donné par  $\kappa = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C)$ , et  $m_{1 \oplus 2}(\emptyset) = 0$ .

La règle de Dempster satisfait

$$plp_{1 \oplus 2}(\omega) = \frac{pl_1(\omega)pl_2(\omega)}{\sum_{\omega' \in \Omega} pl_1(\omega')pl_2(\omega')}, \quad \forall \omega \in \Omega, \quad (2)$$

Autrement dit, l'approximation probabiliste de la combinaison de Dempster de deux FM peut se calculer directement à partir des fonctions de contour.

Pour prendre en compte la fiabilité d'une FM, Shafer [10] a introduit le concept d'affaiblissement. Une FM  $m$  peut être affaiblie à l'aide d'un degré de fiabilité  $\beta = 1 - \alpha \in [0, 1]$  [11] conduisant à une nouvelle FM notée  ${}^\beta m$  définie par :

$${}^\beta m = \beta m + \alpha m_\Omega \quad (3)$$

avec  $m_\Omega$  la fonction d'ignorance totale définie par  $m_\Omega(\Omega) = 1$ . La fonction de contour associée à la FM affaiblie  ${}^\beta m$  est donnée par :

$${}^\beta pl(\omega) = 1 - (1 - pl(\omega))\beta \quad (4)$$

pour tout  $\omega \in \Omega$ , où  $pl$  désigne la fonction de contour de  $m$  [11, 7, 9].

Cette opération a inspiré l'affaiblissement contextuel (CD) dans lequel un degré de fiabilité spécifique  $\beta_\omega \in [0, 1]$  est attribué à chaque élément  $\omega$  de  $\Omega$ . La fonction de masse résultante  ${}^\beta m$  est donnée par :

$${}^\beta m(A) = \sum_{B \subseteq A} m(B) \prod_{\omega_k \in A \setminus B} (1 - \beta_k) \prod_{\omega_k \in \bar{A}} \beta_k, \quad (5)$$

et la fonction de contour associée est définie par :

$${}^\beta pl_{CD}(\omega) = 1 - (1 - pl(\omega))\beta_\omega. \quad (6)$$

### 3 Réseau de neurones crédibiliste à affaiblissement contextuel (CD-ENN)

Dans cette section, nous commençons par faire un rappel d'ENN dans la Section 3.1 avant d'introduire CD-ENN dans la Section 3.2

#### 3.1 ENN

ENN est un réseau de neurones crédibiliste qui à partir des distances entre une instance de test  $x$  dont la classe  $y \in \Omega = \{\omega_1, \dots, \omega_c\}$  est inconnue et des prototypes, construit une FM à propos de  $y$ , grâce à la règle de Dempster. Il est défini avec un ensemble de prototypes  $\{p^i\}_{i=1}^n$ , où chaque prototype  $p^i \in \mathbb{R}^p$  est muni de deux paramètres  $\alpha^i \in ]0, 1[$  et  $\gamma^i > 0$ , ainsi que d'un vecteur de proportions de classes  $u^i = (u_q^i)_{q=1}^c$ , tel que  $u_q^i \in [0, 1]$  et  $\sum_{q=1}^c u_q^i = 1$ . Ainsi, pour une instance de test  $x$  chaque prototype  $p^i$  induit une fonction de masse

$$\begin{aligned} \hat{m}^i(\{\omega_q\}) &= u_q^i s^i, \quad q = 1, \dots, c \\ \hat{m}^i(\Omega) &= 1 - s^i \end{aligned} \quad (7)$$

où  $s^i = \alpha^i \exp(-\gamma^i (d^i)^2)$  avec  $d^i = \|x - p^i\|$  la distance Euclidienne séparant  $x$  du prototype  $p^i$ . La FM  $\hat{m}^i$  associée à chaque prototype  $p^i$  peut aussi être notée vectoriellement par :

$$\hat{m}^i = (\hat{m}^i(\{\omega_1\}), \dots, \hat{m}^i(\{\omega_c\}), \hat{m}^i(\{\Omega\}))^T \quad (8)$$

En combinant les fonctions de masses issues des prototypes par la règle de Dempster, nous obtenons la fonction de masse  $\hat{m}$  suivante :

$$\hat{m} = \bigoplus_{i=1}^n \hat{m}^i \quad (9)$$

La FM  $\hat{m}^i$  (7) induite par chaque prototype  $p^i$  peut être interprétée comme un affaiblissement classique de la fonction de masse  $m^i(\{\omega_q\}) = u_q^i$  pour tout  $q = 1, \dots, c$ , avec un degré de fiabilité égal à  $1 - s^i$ .

Au total, ENN possède  $n(p+c+2)$  paramètres :  $(p^i, u^i, \alpha^i, \gamma^i)_{i=1}^n$ .

#### 3.2 CD-ENN

Dans la section précédente, nous avons vu que dans le classifieur ENN, les fonctions de masses  $\hat{m}^i$  induites par les prototypes peuvent être interprétées comme un affaiblissement classique d'une FM  $m^i$  avec un degré de fiabilité  $1 - s^i$ . En remplaçant cet affaiblissement par CD, on obtient d'après (5), que la fonction de masse induite par chaque prototype  $p^i$  devient :

$$\hat{m}^i(A) = \sum_{\omega_k \in A} m^i(\omega_k) \left( \prod_{\omega_q \in A \setminus \{\omega_k\}} (1 - s_q^i) \times \prod_{\omega_\ell \in \bar{A}} s_\ell^i \right) \quad (10)$$

où  $m^i(\{\omega_q\}) = u_q^i$ . Les coefficients  $s_q^i$  sont définis par :

$$s_q^i = \alpha^i \exp(-\gamma^i d^i{}^2) \quad q = 1, \dots, c \quad (11)$$

Chaque  $s_q^i$  peut être interprété comme le degré de confiance qu'apporte le prototype  $p^i$ , situé à distance  $d^i$  de  $x$ , vis-à-vis de la classe  $\omega_q$ .

L'expression de la fonction de contour associée à la masse induite par chaque prototype (10) peut être donnée grâce à l'équation (6), par :

$$\hat{pl}^i(\omega_q) = 1 - s_q^i + s_q^i u_q^i, \quad q = 1, \dots, c \quad (12)$$

Notons enfin, que contrairement à ENN qui donne comme possibles éléments focaux que les singletons et l'univers  $\Omega$ , CD-ENN peut affecter une masse positive à tous les sous-ensembles de  $\Omega$ .

Le modèle CD-ENN partage des paramètres communs avec ENN, à savoir l'ensemble des prototypes  $\{p^i\}_{i=1}^n$  munis de leurs vecteurs de proportions de classes  $\{u^i\}_{i=1}^n$ , et de leurs paramètres alpha  $\{\alpha^i\}_{i=1}^n$ . En revanche, il diffère au niveau des paramètres  $\{\gamma^i\}_{i=1}^n$ . En effet, alors que ENN est défini avec des paramètres scalaires  $\gamma^i$ , (c'est-à-dire un paramètre par prototype), CD-ENN est lui défini avec des vecteurs  $\gamma^i = (\gamma_q^i)_{q=1}^c$  (c'est-à-dire un paramètre par prototype et par classe). Par conséquent, CD-ENN dispose de  $n(p + 2c + 1)$  paramètres  $(p^i, u^i, \alpha^i, \gamma^i)_{i=1}^n$ .

## 4 Premières expérimentations

Nous montrons dans cette partie des expérimentations sur deux jeux de données générées où CD-ENN améliore ENN.

### 4.1 Jeux de données générés

Un **jeu de données à trois classes** (600 échantillons chacune) est généré :

- *Classe 1* :  $\mathcal{N}((300, 310)^T, 5000I)$ ,
- *Classe 2* : Points uniformément répartis sur le cercle centré en  $(300, 302)^T$  de rayon 200,
- *Classe 3* :  $\mathcal{N}((420, 280)^T, \Sigma_3)$ ,  $\Sigma_3 = \begin{bmatrix} 20000 & 10000 \\ 10000 & 10000 \end{bmatrix}$

Un second **jeu de données à 5 classes** est généré avec une classe 1 de densité  $\mathcal{N}((0, 0)^T, 0.2I)$ , une classe 2 de densité  $\mathcal{N}((2, 2)^T, 0.5 \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix})$ , une classe 3 avec des points uniformément répartis sur un cercle de rayon moyen 1, bruités selon  $\mathcal{N}(1, 0.05^2)$ . une classe 4 avec des points uniformément répartis sur un cercle de rayon moyen 2, bruités selon  $\mathcal{N}(2, 0.1^2)$ , et enfin une classe 5 avec

une spirale paramétrée par  $\theta \in [0, 4\pi]$ ,  $r = 0.1\theta$ ,  $(x, y) = (r \cos \theta, r \sin \theta)$ .

Un échantillon de chacun de ces jeux de données est illustré sur la Figure 1.

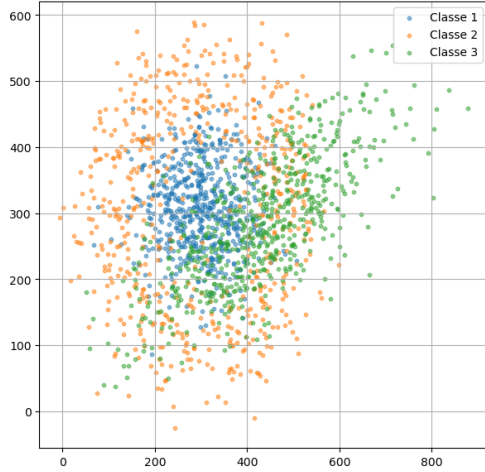
### 4.2 Méthodologie

Les réseaux CD-ENN et ENN sont entraînés en minimisant l'entropie croisée de l'approximation probabiliste de  $\hat{m}$  définie par (9) avec  $\hat{m}^i$  définie par (7) pour ENN et  $\hat{m}^i$  définie par (10) pour CD-ENN, pour tout  $i = 1, \dots, n$ . Grâce à la propriété de la règle de Dempster relative aux produits des fonctions de contour (2) et à l'expression des fonctions de contour (12) issues des prototypes, la combinaison par la règle de Dempster des fonctions de contour issues des prototypes se calcule pour CD-ENN en un temps linéaire par rapport au nombre  $c$  de classes (comme pour ENN).

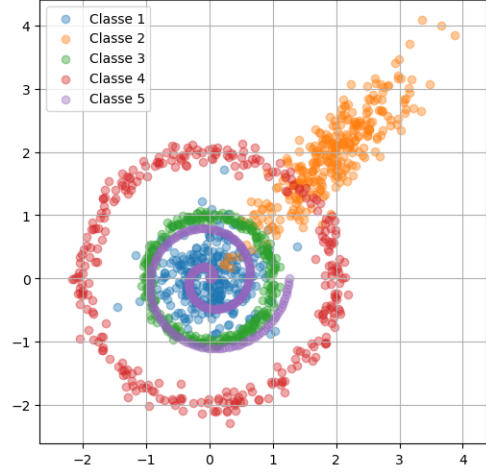
Les prototypes sont initialisés par un clustering k-means en prenant deux fois plus de prototypes que de classes présentes dans le jeu ( $n = 2c$ ). Chaque degré d'appartenance initial  $u_q^i$  du prototype  $p^i$  est déterminé comme la proportion d'exemples d'apprentissage de la classe  $\omega_q$  dans le cluster dont le centre est  $p^i$ . Pour le classifieur CD-ENN, les paramètres restants sont initialisés pour tout prototype  $p^i$  et classe  $\omega_q$  par  $\gamma_q^i = 0.1$  et  $\alpha^i = 0.5$ . De même pour le classifieur ENN, les paramètres restants sont initialisés pour tout prototype  $p^i$  par  $\gamma^i = 0.1$  et  $\alpha^i = 0.5$ .

Pour les expérimentations, nous avons utilisé une division stratifiée aléatoire (en anglais *Stratified Shuffle Split*) afin de préserver la proportion des classes dans chaque partition en prenant 80% des données pour l'entraînement et 20% pour la phase de test.

Puisque pour la vraie classe  $y$  d'une instance de vecteur  $x$ , on dispose non plus d'une unique probabilité mais de deux quantités à savoir le degré de croyance  $Bel_x(\{y\})$  et le degré de plausibilité  $Pl_x(\{y\})$ , chaque modèle peut être évalué en utilisant la fonction de perte *Generalized*



(a) Jeu de données à 3 classes



(b) Jeu de données à 5 classes

FIGURE 1 – Échantillons des jeux de données générées

zed Negative Log-Likelihood (GNLL) proposée par Denœux [6] et définie par :

$$\mathcal{L}(y, m_x) = -\frac{1}{2} \ln Bel_x(\{y\}) - \frac{1}{2} \ln Pl_x(\{y\}). \quad (13)$$

Aussi utilisée dans [8], cette fonction de perte est d'autant plus faible que les degrés de croyance et de plausibilité sont élevés pour la vraie classe. En outre, dans le cas d'une FM  $m_x$  bayésienne, c'est-à-dire où  $m_x$  est une distribution de probabilité  $P_x$ , nous avons  $Bel_x(\{y\}) = Pl_x(\{y\}) = P_x(\{y\})$ , et cette fonction de perte se réduit à l'entropie croisée :  $\mathcal{L}(y, m_x) = -\ln P_x(\{y\})$ .

### 4.3 Résultats

En répétant 50 fois cette procédure, nous obtenons les résultats de pertes GNLL illustrés sur la Figure 2 montrant une amélioration de ce score pour CD-ENN.

Nous avons également rencontré des jeux de données générées pour lesquels ENN donne de meilleurs résultats que CD-ENN au sens de la fonction de perte GNLL. Il semble que ce serait le cas si les données sont séparées, par exemple en prenant 3 îlots distincts de points. Cela reste à approfondir.

## 5 Conclusion

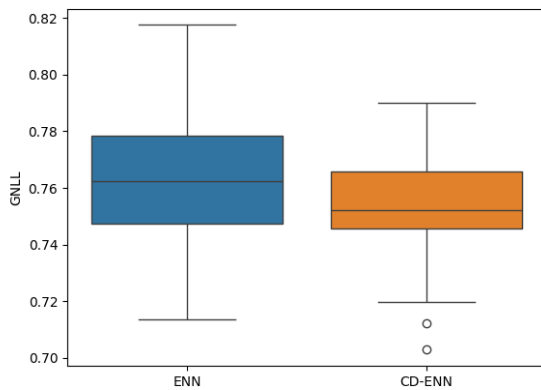
Dans cet article, une première étude exploratoire d'une extension du modèle ENN de Denœux est proposée en remplaçant l'affaiblissement classique par l'affaiblissement contextuel. Cette approche introduit un degré de fiabilité par prototype et par classe. Des premières expérimentations sur des données générées montrent un intérêt à cette approche. Ces expérimentations seront étoffées par exemple sur des jeux de données UCI. Notre objectif est également de montrer l'intérêt de cette approche en remplaçant la couche softmax d'un réseau de neurones profonds tout en étant capable de faire un apprentissage de bout en bout.

### Remerciements :

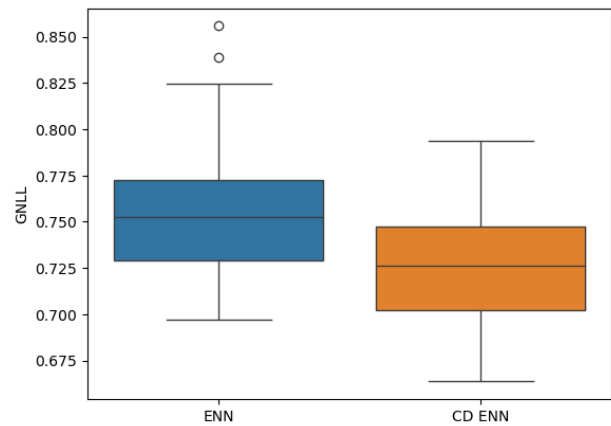
Ces travaux sont cofinancés par la région Hauts-de-France et l'université d'Artois dans le cadre de la thèse de M. Serigne Mamadou Diène.

### Références

- [1] A. P. DEMPSTER. "Upper and lower probabilities induced by a multiple valued mapping". In : *The Annals of Mathematical Statistics* 38 (1967), p. 325-339.
- [2] T. DENŒUX. "A k-nearest neighbor classification rule based on Dempster-Shafer theory". In : *IEEE Transactions on Sys-*



(a) Jeu de données à 3 classes



(b) Jeu de données à 5 classes

FIGURE 2 – Pertes GNLL ( $\downarrow$ ) obtenues sur les deux jeux de données.

- tems, Man, and Cybernetics* 25.5 (1995), p. 804-813.
- [3] T. DENÈUX. “A neural network classifier based on Dempster–Shafer theory”. In : *IEEE Transactions on Systems, Man, and Cybernetics A* 30.2 (2000), p. 131-150.
- [4] T. DENÈUX. “Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence”. In : *Artificial Intelligence* 172 (2008), p. 234-264.
- [5] T. DENÈUX. “Logistic regression, neural networks and Dempster–Shafer theory : A new perspective”. In : *Knowledge-Based Systems* 176 (2019), p. 54-67.
- [6] T. DENÈUX. “Quantifying Prediction Uncertainty in Regression using Random Fuzzy Sets : the ENNreg model”. In : *IEEE Transactions on Fuzzy Systems* 31.10 (2023), p. 3690-3699.
- [7] D. MERCIER, B. QUOST et T. DENÈUX. “Refined Modeling of Sensor Reliability in the Belief Function Framework Using Contextual Discounting”. In : *Information Fusion* 9.2 (2008), p. 246-258.
- [8] F. PICHON, S. DIÈNE, T. DENÈUX, S. RAMEL et D. MERCIER. “r-ERBFN : an Extension of the Evidential RBFN Accounting for the Dependence Between Positive and Negative Evidence”. In : *Scalable Uncertainty Management*. Sous la dir. de S. DESTERCKE, M. VANINA MARTINEZ et G. SANFILIPPO. T. 15350. Lecture Notes in Artificial Intelligence. Palermo, Italy : Springer, 2024.
- [9] F. PICHON, D. MERCIER, E. LEFÈVRE et F. DELMOTTE. “Proposition and learning of some belief function contextual correction mechanisms”. In : *International Journal of Approximate Reasoning* 72 (2016), p. 4-42.
- [10] G. SHAFER. *A Mathematical Theory of Evidence*. Princeton, N.J. : Princeton University Press, 1976.
- [11] P. SMETS. “Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem”. In : *International Journal of Approximate Reasoning* 9.1 (1993), p. 1-35.
- [12] Z. TONG, P. XU et T. DENÈUX. “An evidential classifier based on Dempster–Shafer theory and deep learning”. In : *Neurocomputing* 450 (2021), p. 275-293.