

Utilisation de corrections dans des réseaux de neurones profonds crédibilistes à couches denses dans une approche de fusion multivue

Use of corrections in evidential dense-layer deep neural networks in a multi-view fusion approach

S.M. Diène¹

S. Ramel¹

F. Pichon¹

D. Mercier¹

¹ 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, dans un cadre de fusion multivue, l'utilisation de différentes corrections contextuelles dans des réseaux de neurones profonds crédibilistes à couches denses (perceptrons multicouches), c'est-à-dire des réseaux où la dernière couche softmax a été remplacée par une couche de Dempster-Shafer pour obtenir une fonction de croyance à la place d'une probabilité et où a été également ajoutée à la suite une couche de correction, puis une couche de fusion et enfin une de décision. Des premiers résultats obtenus sur des données expérimentales montrent l'intérêt de cette approche.

Mots-clés :

Réseaux de neurones profonds, Couche de Dempster-Shafer, Corrections

Abstract:

In this article, we explore, within a multi-view fusion framework, the use of different contextual corrections in evidential deep neural networks with dense layers (multilayer perceptrons), i.e. networks where the last softmax layer has been replaced by a Dempster-Shafer layer to obtain a belief function instead of a probability, and where a correction layer has also been added subsequently, followed by a fusion layer and finally a decision layer. Initial results obtained from experimental data show the interest of this approach.

Keywords:

Deep neural networks, Dempster-Shafer layer, Corrections

1 Introduction

Des travaux récents [13, 7] ont montré l'intérêt d'enrichir des réseaux de neurones profonds probabilistes en remplaçant la couche finale de décision par une couche de Dempster-Shafer [3, 13, 7] afin d'obtenir en sortie une fonction de

croyance [2, 12, 3] donnant ainsi une information plus riche, mieux représentative des incertitudes et imprécisions en sortie.

D'un autre côté, pour corriger/affiner une fonction de croyance à partir d'un ensemble de données étiquetées, il existe dans la théorie de Dempster-Shafer une opération d'affaiblissement [12, 11] ou des corrections plus fines telles que les corrections contextuelles [8, 10] parmi lesquelles nous trouvons l'affaiblissement contextuel (noté CD, *Contextual Discounting*), permettant d'affaiblir une information et qui généralise l'affaiblissement, ou le renforcement contextuel (noté CR, *Contextual Reinforcement*), pouvant renforcer une information, ou le reniement contextuel (notée CN, *Contextual Negating*), capable de nier ce qu'une source indique.

Huang et al. [7] ont récemment exploré l'idée d'ajouter une couche de correction par affaiblissement contextuel à la suite de la couche de Dempster-Shafer pour proposer une approche de segmentation d'image par fusion multivue de différents encodeurs-décodeurs associés à différentes modalités ou vues d'une image.

Dans cet article, nous proposons d'explorer plus loin cette approche de fusion multivue en étudiant également ce que peuvent apporter d'autres mécanismes de corrections contex-

tuelles, comme le renforcement et le reniement, et avec des réseaux classiques comme les perceptrons multicouches.

Cet article est organisé de la manière suivante. Les notations et notions utilisées du domaine des fonctions de croyance dont les corrections sont exposées dans la Section 2. Les réseaux crédibilistes utilisant différentes corrections considérés sont décrits dans la Section 3. Des expériences sur des données réelles montrant l'intérêt de ces approches sont alors présentées dans la Section 4. Enfin, une conclusion et des perspectives sont données dans la Section 5.

2 Fonctions de croyance et corrections

Les principaux concepts utilisés dans cet article de la théorie des fonctions de croyance sont présentés ici. Des détails sur cette théorie peuvent être trouvés par exemple dans [12, 11, 4, 5].

Une information au regard de la réponse à une question Q d'intérêt dont la réponse est dans un ensemble fini Ω composé de K éléments $\omega_1, \dots, \omega_K$, peut être représentée par une fonction de masse (FM) m définie de 2^Ω dans $[0, 1]$, t.q. $\sum_{A \subseteq \Omega} m(A) = 1$.

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

Une FM m est en correspondance biunivoque avec une fonction de plausibilité Pl définies pour tout $A \subseteq \Omega$ par $Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$.

La *fonction contour* pl correspond à la restriction de la fonction de plausibilité aux singletons de Ω : $pl(\omega) = Pl(\{\omega\})$, pour tout $\omega \in \Omega$.

Deux FM fiables et indépendantes m_1 et m_2 définies sur le même univers Ω peuvent être combinées en utilisant la règle de Dempster définie pour tout $\emptyset \neq A \subseteq \Omega$ par : $(m_1 \oplus m_2)(A) = m_{1 \oplus 2}(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) \cdot$

$m_2(C)$, avec $\kappa = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C)$, et $m_{1 \oplus 2}(\emptyset) = 0$.

Notons également que la fonction de contour de la règle de Dempster s'obtient aisément par

$$(pl_1 \oplus pl_2)(\omega) = \frac{pl_1(\omega) pl_2(\omega)}{1 - \kappa}, \quad \forall \omega \in \Omega \quad (1)$$

Shafer [12] introduit également la notion d'affaiblissement ${}^\beta m$ d'une fonction de masse m pour tenir compte de son degré de fiabilité $\beta = 1 - \alpha \in [0, 1]$ [11]. Cet affaiblissement est donné par ${}^\beta m = \beta m + \alpha m_\Omega$, avec m_Ω l'ignorance totale définie par $m_\Omega(\Omega) = 1$. La fonction de contour associée à cette FM affaiblie ${}^\beta m$ vérifie pour tout $\omega \in \Omega$, ${}^\beta pl(\omega) = 1 - (1 - pl(\omega))\beta$, avec pl la fonction de contour de m [11, 8, 10].

Cet affaiblissement se généralise en l'affaiblissement contextuel (CD) avec K paramètres $\beta_\omega \in [0, 1]$ pour tout $\omega \in \Omega$ [8, 10], qui est défini par :

$${}^\beta pl_{CD}(\omega) = 1 - (1 - pl(\omega))\beta_\omega \quad (2)$$

Il est également possible de prendre en compte d'autres notions de corrections de biais d'une source pour définir le renforcement contextuel (CR) et le reniement contextuel (CN) d'une source [10], avec également K paramètres $\beta_\omega \in [0, 1]$ pour tout $\omega \in \Omega$. Ces deux corrections sont respectivement définies par :

$${}^\beta pl_{CR}(\omega) = pl(\omega)\beta_\omega \quad (3)$$

et

$${}^\beta pl_{CN}(\omega) = 0.5 + (pl(\omega) - 0.5)(2\beta_\omega - 1) \quad (4)$$

Seules les versions exprimées en termes de fonctions de contour et sur K paramètres, ont été données ici pour CD, CR et CN. Elles suffisent sur ces travaux. Il existe des modèles plus généraux plus complexes à employer [10].

Enfin, il est possible de transformer une FM en une mesure de probabilité en normalisant sa fonction de contour :

$$plp(\omega) = \frac{pl(\omega)}{\sum_{\theta \in \Omega} pl(\theta)}, \quad \forall \omega \in \Omega \quad (5)$$

Une telle transformation présente de bonnes propriétés [17, 1] dont l'avantage de pouvoir calculer rapidement, avec une faible complexité, la probabilité finale d'un élément après une fusion par la règle de Dempster :

$$plp_{m_1 \oplus m_2}(\omega) \propto plp_{m_1}(\omega) plp_{m_2}(\omega), \quad \forall \omega \in \Omega, \quad (6)$$

d'après les équations (1) et (5).

3 Réseaux crédibilistes considérés

Comme Denœux l'a montré dans [3], il est possible de définir un classifieur crédibiliste sous la forme d'un réseau de neurones (classifieur appelé *ENN* pour *Evidential Neural Network*), basé sur des distances à des prototypes pour définir des fonctions de croyance combinées par la règles de Dempster. Cette *couche de Dempster-Shafer* peut alors être utilisée en remplacement de la couche softmax finale d'un réseau de neurones [13], avec également la possibilité d'y ajouter une correction [7].

Dans cet article, nous étudions une première idée illustrée sur la Figure 1. Nous considérons plusieurs réseaux de neurones crédibilistes, c'est-à-dire des réseaux de neurones avec ajout d'une couche de Dempster-Shafer, à chacun desquels une couche de correction (CD, CR, ou CN) a été ajoutée, ceci dans un contexte de fusion d'informations.

L'idée de considérer une fusion d'informations en sortie de plusieurs classifieurs crédibilistes est de mettre en avant l'intérêt (et de profiter) d'avoir une fonction de croyance en sortie d'un classifieur plutôt qu'une mesure de probabilité.

4 Premières expérimentations

Dans cet article, nous considérons les 5 modèles suivants :

1. **PDNN** (*Probabilistic Deep Neural Networks*) : il s'agit ici de perceptrons multicouches à deux couches cachées, chacun avec une couche finale softmax et chacun associé à chaque vue. La couche

fusion est une multiplication des probabilités [14, 15].

2. **EDNN** (*Evidential Deep Neural Networks*) : idem que le modèle PDNN sauf que la couche softmax de chaque perceptron est remplacée par une couche de Dempster-Shafer, suivie par une transformation m vers pl et une utilisation de la propriété (6) pour fusionner par la règle de Dempster (produit des fonctions de contour) et prendre une décision par maximum de plausibilité transformée plp .
3. **CDEDNN** (*Contextually Discounted Evidential Deep Neural Networks*) : idem que le modèle EDNN sauf qu'on ajoute une couche de correction CD après chaque couche de Dempster-Shafer (cf Figure 1).
4. **CREDDNN** (*Contextually Reinforced Evidential Deep Neural Networks*) : idem que le modèle CDEDNN sauf que les corrections CD sont remplacées par des corrections CR.
5. **CNEDNN** (*Contextually Negated Evidential Deep Neural Networks*) : idem que le modèle CREDDNN sauf que les corrections CR sont remplacées par des corrections CN.

Nous utilisons le jeu de données multivues nommé *Multiple Features* [6, 16] composé de 2000 écritures manuscrites des 10 chiffres (0 à 9) avec 200 échantillons pour chaque chiffre et 6 vues de caractéristiques disponibles : coefficients de Fourier, corrélations de profil, coefficients de Karhunen-Love, moyennes de pixels dans des fenêtres de 2 x 3, moments de Zernike et 6 caractéristiques morphologiques. Ce jeu de données est notamment accessible via le module Python *mvlearn* [9].

L'apprentissage des réseaux de neurones est effectuée en utilisant l'entropie croisée comme fonction de coût (approche classique de classification), la transformation de la fonction de contour en probabilité en sortie des réseaux crédibilistes via plp , Équation 5, l'algorithme

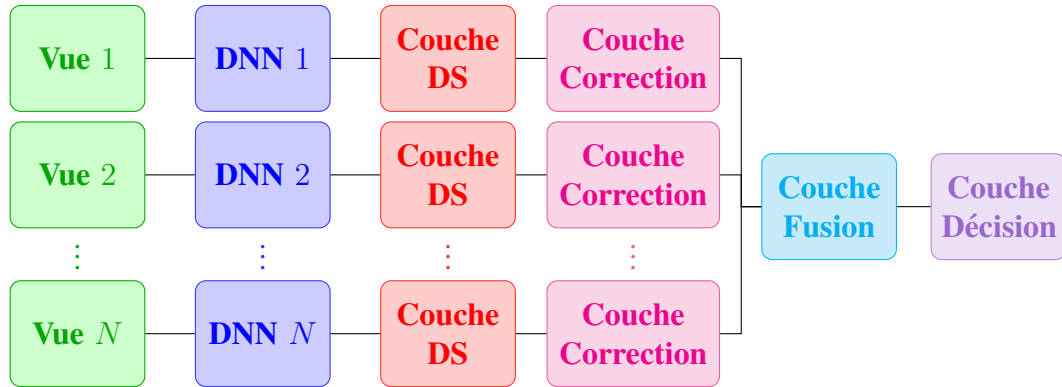


FIGURE 1 – Modèles proposés de réseaux crédibilistes avec corrections (CD, CR ou CN) et fusion multivue

d’optimisation Adam, et la mesure de performance du taux de bonnes reconnaissances (*accuracy*). Sauf mention contraire, chaque apprentissage décrit ci-après se fait sur 100 époques, avec 25 comme taille de batch. L’optimisation de chaque paramètre β_ω , contraint à appartenir à l’intervalle $[0, 1]$, est reformulée en problème d’optimisation sans contrainte, en utilisant la reparamétrisation $\gamma = \log(\beta_\omega/1 - \beta_\omega)$, où le nouveau paramètre γ appartient alors à l’ensemble des réels et est initialisé via une distribution normale.

Pour obtenir nos résultats, nous répétons 10 fois la procédure d’apprentissage suivante :

- Chaque réseau du PDNN est appris individuellement, puis l’ensemble PDNN (réseaux avec fusion) est appris à partir des poids précédents.
- Puis chaque réseau crédibiliste de EDNN est appris individuellement avec 10 prototypes en fixant d’abord les poids avant la couche de Dempster-Shafer, à ceux appris dans le réseau correspondant du PDNN et en optimisant les poids de la couche de Dempster-Shafer sur 10 époques, puis en dégelant les poids préalablement fixés (*fine-tuning*) et en optimisant la totalité des poids du réseau EDNN complet sur 100 époques afin de permettre un réglage fin de ces derniers.

- CDEDNN, CREDNN et CNEDNN sont ensuite appris de la même manière qu’EDNN outre le fait que ce sont les poids d’EDNN qui sont réutilisés au lieu de ceux du modèle PDNN.

Dans cette procédure, le même lot de données choisi aléatoirement pour l’apprentissage est donc utilisé pour l’apprentissage de tous les réseaux.

En prenant aléatoirement 50% des données pour l’apprentissage, nous obtenons les performances données dans le Tableau 1a, et en considérant 80% des données pour l’apprentissage, nous obtenons celles indiquées dans le Tableau 1b.

Bien que modestes, on peut quand même voir des améliorations des taux de bonnes reconnaissances (*accuracy*), et ainsi motiver une exploration plus importante de ces modèles de réseaux crédibilistes enrichis de couches de correction. Des pistes de recherches sont mentionnées dans la section suivante.

5 Conclusion et perspectives

Dans cet article, une approche générale connexionniste de fusion crédibiliste avec corrections a été présentée. Elle consiste à ajouter à des EDNN une couche de correction avant de les fusionner. Sur une base de tests à 6 vues,

TABLEAU 1 – Performances moyennes sur 10 répétitions de la partition aléatoire apprentissage/test et l'apprentissage des réseaux présentés sur le jeu de données *Multiple Features* [6, 16] (écarts types indiqués entre parenthèses, meilleure performance en gras)

(a) Avec 50% des données pour l'apprentissage

Réseau	Accuracy
PDNN	0.9626 (0.0063)
EDNN	0.9579 (0.0065)
CDEDNN	0.9654 (0.0062)
CREDNN	0.9644 (0.0075)
CNEDNN	0.9590 (0.0085)

(b) Avec 80% des données pour l'apprentissage

Réseau	Accuracy
PDNN	0.9849 (0.0039)
EDNN	0.9864 (0.0025)
CDEDNN	0.9877 (0.0020)
CREDNN	0.9889 (0.0027)
CNEDNN	0.9871 (0.0019)

nous avons vu une amélioration du taux de bonnes reconnaissances (*accuracy*) avec 50% ou 80% des données en apprentissage. Il s'agit de premiers résultats qui doivent être étoffés sur plus de bases de tests.

Plusieurs idées de perspectives vont être explorées. Pour l'instant, un même modèle de correction est commun à toutes les vues, il serait possible d'envisager la meilleure correction pour chaque vue, voire aussi d'explorer des corrections successives. L'appréciation de la "meilleure" qualité de sortie pourrait aussi être mise en avant par d'autres mesures de performances notamment en classification prudente.

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] B.R. Cobb, P.P. Shenoy, On the plausibility transformation method for translating belief function mo-

dels to probability models, *International Journal of Approximate Reasoning*, 41(3) :314–330, 2006.

[2] A.P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *The Annals of Mathematical Statistics*, 38 :325–339, 1967.

[3] T. Denœux. A neural network classifier based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics A*, 30(2) :131–150, 2000.

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

[5] T. Denœux. Logistic regression, neural networks and Dempster-Shafer theory : A new perspective. *Knowledge-Based Systems*. 176 :54–67, 2019.

[6] R. Duin. Multiple Features. *UCI Machine Learning Repository*. <https://doi.org/10.24432/C5HC70>.

[7] L. Huang, T. Denœux, P. Vera, S. Ruan. Evidence Fusion with Contextual Discounting for Multimodality Medical Image Segmentation. In L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, S. Li (Eds), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. *Lecture Notes in Computer Science*, vol 13435. Springer, Cham, 2022.

[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] R. Perry, G. Mischler, R. Guo, T. Lee, A. Chang, A. Koul, C. Franz, H. Richard, I. Carmichael, P. Ablin, A. Gramfort, J.-T. Vogelstein. *mvlearn : Multiview Machine Learning in Python*. *Journal of Machine Learning Research*, 22(109) :1–7, 2021.

[10] 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.

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

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

[13] Z. Tong, P. Xu, T. Denœux. An evidential classifier based on Dempster-Shafer theory and deep learning. *Neurocomputing*. 450 :275–293, 2021.

[14] 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.

[15] P. Xu, F. Davoine, J.B. Bordes, H. Zhao, T. Denœux. Multimodal Information Fusion for Urban Scene Understanding. *Machine Vision and Applications*, 27(3) :331–349, 2016.

[16] M. Van Breukelen, R.P.W. Duin, D.M.J. Tax, J.E. Den Hartog. Handwritten digit recognition by combined classifiers, *Kybernetika*, 34(4) :381–386, 1998.

[17] F. Voorbraak. A computationally efficient approximation of Dempster-Shafer theory, *International Journal of Man-Machine Studies*, 30 :525–536, 1989.