On improving a group of evidential sources with different contextual corrections

Siti Mutmainah[1] [0000-0001-7747-435X], Samir Hachour[2] [0000-0001-6483-6833], Frédéric Pichon[2] [0000-0002-9905-5882], and David Mercier[2] [0000-0002-7511-7874]

[1] UIN Sunan Kalijaga, Yogyakarta, Indonesia
`siti.mutmainah@uin-suka.ac.id`
[2] Univ. Artois, UR 3926 LGI2A, Béthune, F-62400, France
`{firstname.lastname}@univ-artois.fr`

Abstract. In this paper, we investigate the interest of learning a group of evidential sources using contextual corrections, which is equivalent to directly learning an optimized conjunctive combination instead of optimizing each source individually. Several experiments on synthetic and real UCI data demonstrates the interest of the approach.

Keywords: Information Fusion · Belief functions · Group of sources · Contextual corrections · Optimization.

1 Introduction

Information fusion [1, 11] allows one, by combining different heterogeneous sources of information, to obtain a better understanding (possibly more complete, more precise) of the situation under evaluation.

The Dempster-Shafer theory of belief functions [18, 2, 17], being able to represent the imprecision and uncertainty of a piece of information, is an interesting and already widely used framework for modeling a fusion scheme [9, 16]. One classical evidential fusion scheme consists in modeling the individual outputs of the sources as finely as possible to make independent and reliable pieces information so that they can be combined using the conjunctive rule of combination (meaning the unnormalized Dempster's rule). The reliability of the outputs of the sources can be ensured using the discounting operation [18, 15, 15] or more refined corrections such that contextual corrections [13, 14]. For instance, we can use the contextual discounting (CD), allowing one to weaken a piece of information and which generalizes the discounting, or the contextual reinforcement (CR), which can reinforce the output of a source, or the contextual negating (CN), able to negate what a source indicates.

In the discounting operation [18], the reliability of the source, providing a mass function m, is taken into account using a real $\beta \in [0, 1]$ quantifying the degree of belief in the fact that the source is reliable, and the corrected mass function is denoted by ${}^{\beta}m$. In the contextual correction mechanisms (CD, CR and CN), the imperfection of the source, its bias in a broad sense, is modeled using a vector $\beta \in [0,1]^C$, with $C \leq 2^K$ and K the number of elements in the universe (more specific details can be found in [14]). The resulting corrected mass function is also denoted by ${}^{\beta}m$ for simplicity.

If moreover, a learning set composed of the outputs of a source, expressed in the form of mass functions, are available regarding the classes of n objects o_i, $i \in \{1, \ldots, n\}$ the true class (belonging to the universe) of each object being known, then it is possible [10, 13] to find optimal parameters β, *i.e.*, to learn the parameters β minimizing a discrepancy measure between the corrected outputs and the ground truths.

This classical information fusion scheme is illustrated in Figure 1.

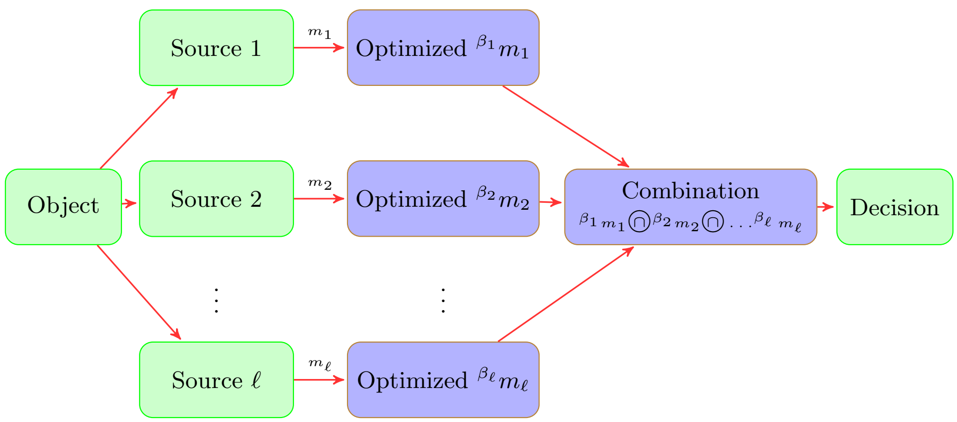

Fig. 1. Fusion scheme using individual corrections (Scheme 1).

Another idea, illustrated in Figure 2, consists in learning directly an optimized conjunctive combination instead of optimizing each source individually. This idea has been mentioned in [10] for the discounting operation and in [13] for a particular CD.

In this paper, we use classifiers as sources of information, and we explore this idea of optimizing directly the performance of the combination using possibly different corrections among CD, CR and CN.

This paper is organized as follows. The notations and evidential concepts used are recalled in Section 2. The learning of contextual corrections for a group of evidential classifiers is presented in Section 3. Experiments on synthetic and real data demonstrating the interest of the approach are exposed in Section 4. Finally, a conclusion is given in Section 5.

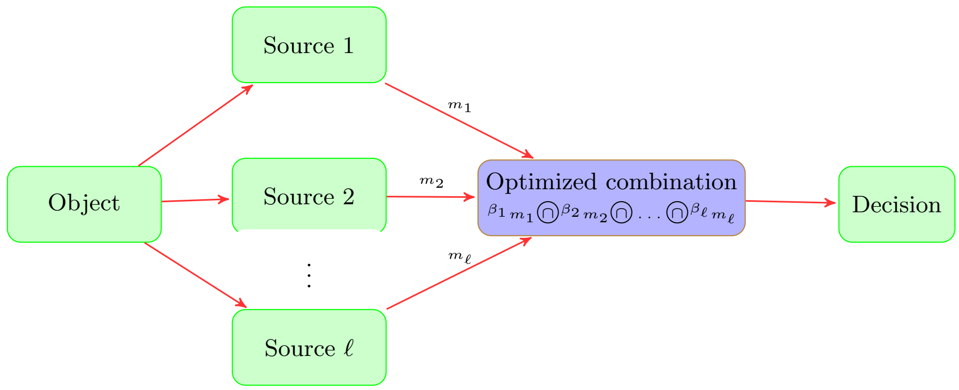

Fig. 2. Fusion scheme using global corrections (Scheme 2).

2 Belief functions: notations and concepts used

2.1 Basic concepts

Basic concepts are briefly recalled. Details of the theory can be found for example in [18, 15, 5].

The universe Ω, a finite set, is composed of K elements $\omega_1, \ldots, \omega_K$. We consider a question of interest Q whose answer lies in Ω. A piece of information regarding this answer can be represented by a mass function (MF) m defined from 2^Ω to $[0,1]$ verifying s.t. $\sum_{A \subseteq \Omega} m(A) = 1$. The real $m(A)$ represents the part of belief allocated to the fact that the true searched value belongs to A and nothing more. A subset $A \subseteq \Omega$ s.t. $m(A) > 0$ is called a focal element of m. A categorical MF has only one focal element $A \subseteq \Omega$ and is denoted by m_A. We then have $m_A(A) = 1$. In particular, m_Ω represents the total ignorance.

A MF m is in one-to-one correspondence with a belief function Bel and a plausibility function Pl respectively defined for all $A \subseteq \Omega$ by:

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \ , \tag{1}$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) = Bel(\Omega) - Bel(\overline{A}) \tag{2}$$

with $\overline{A} = \Omega \setminus A$.

The *contour function pl* corresponds to the restriction of the plausibility function to the singletons of Ω, it is defined for all $\omega \in \Omega$ by $pl(\omega) = Pl(\{\omega\})$.

Two reliable and independent MFs m_1 and m_2 defined on the same universe Ω can be combined using the conjunctive rule of combination (CRC) (or unnormalized Dempster's rule) defined by

$$(m_1 \textcircled{\cap} m_2)(A) = m_{1 \textcircled{\cap} 2}(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \quad \forall A \subseteq \Omega \ . \tag{3}$$

2.2 Corrections

A source providing a MF m and only reliable at a degree $\beta = 1 - \alpha \in [0,1]$ can be discounted using the following operation

$$\begin{aligned} {}^{\beta}m &= \beta\, m + \alpha\, m_{\Omega} \\ &= \begin{cases} A \mapsto \beta\, m(A) & \forall A \subset \Omega \\ \Omega \mapsto \beta\, m(\Omega) + \alpha \end{cases} \end{aligned} \tag{4}$$

The contour function associated with the discounted MF ${}^{\beta}m$ (4) verifies for all $\omega \in \Omega$, ${}^{\beta}pl(\omega) = 1 - (1 - pl(\omega))\beta$, with pl the contour function of m (Details can be found for example in [15, 13, 14]).

In Table 1, we summarize the contour functions of the contextual discounting (CD), contextual reinforcement (CR) and contextual negating (CN) of a MF m that can be obtained by a specific choice of $C = K$ parameters $\beta_w \in [0,1]$, for each contextual corrections; the reasons for limiting ourselves to $C = K$ parameters and the definitions of these K parameters for each contextual corrections can be found in [14, Section 8].

Table 1. Contour functions of each contextual correction of a MF m given for any $\omega \in \Omega$. Each parameter β_ω may vary in $[0,1]$.

Corrections	Contour functions
CD	${}^{\beta}pl(\omega) = 1 - (1 - pl(\omega))\beta_\omega$
CR	${}^{\beta}pl(\omega) = pl(\omega)\beta_\omega$
CN	${}^{\beta}pl(\omega) = 0.5 + (pl(\omega) - 0.5)(2\beta_\omega - 1)$

As recalled in the introduction, if for a source we have a learning set containing its outputs, meaning MF $m\{o_i\}$, regarding the classes of n objects o_i, $i \in \{1, \ldots, n\}$, the true classes are known, we can then compute the CD, CR and CN parameters β optimizing the following measure of discrepancy between the corrected outputs and the true classes of the objects

$$E_{pl}(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} ({}^{\beta}pl\{o_i\}(\{\omega_k\}) - \delta_{i,k})^2 \ , \tag{5}$$

where ${}^{\beta}pl\{o_i\}$ is the contour function regarding the class of the object o_i corrected with a vector $\beta = (\beta_\omega \in [0,1], \omega \in \Omega)$ and $\delta_{i,k}$ is the indicator function of the truths of the objects o_i, $i \in \{1, \ldots, n\}$, meaning $\delta_{i,k} = 1$ if the class of the object o_i is ω_k, otherwise $\delta_{i,k} = 0$.

The measure E_{pl} yields, for each correction (CD, CR, and CN), a constrained linear least-squares optimization problem which can be efficiently solved.

3 Learning a group of evidential sources

When several sources are available, instead of learning the best correction parameters individually for each source knowing that these adjusted MFs are going to be next combined, it is possible to directly optimize the combination of the adjusted MFs.

With ℓ sources to be combined, ℓ vectors $\beta_1, \ldots, \beta_\ell$, each one associated with either CD or CR or CN, can be obtained by minimizing the following measure

$$E_{pl}(\beta_1, \ldots, \beta_\ell) = \sum_{i=1}^{n} \sum_{k=1}^{K} ({}^{\beta_1}pl_1\{o_i\}(\{\omega_k\}) \times \ldots \times^{\beta_\ell} pl_\ell\{o_i\}(\{\omega_k\}) - \delta_{i,k})^2 \quad (6)$$

Indeed, after the conjunctive combination, the plausibility of each singleton is equal to the product of the plausibilities given by the ℓ sources to this singleton.

Optimizing (5) for each classifier or (6) is not the same thing as a classifier can be used in a different manner if it is used alone or through a collective.

One drawback, however, of this approach, is that the optimization of (6) is no more a linear least-squares optimization problem, it can be minimized using a standard constrained nonlinear optimization procedure reaching to a possible local minimum.

Another critical point concerns the number of optimizations to undertake in each scenario. With three possible mechanisms (CD, CR and CN), which can be applied on each source, and ℓ sources, we have for the first scheme using individual corrections (cf Figure 1) $3 \times \ell$ possible corrections to test, while for this second scheme optimizing the combination (cf Figure 2), we have 3^ℓ possible corrections to test.

As an example, let us consider the case of two sources S_1 and S_2 ($\ell = 2$). For the individual optimizations, we have for each source three optimisations to undertake, using (5), to know what correction between CD, CR and CN to keep for each source, and thus finally 6 optimisations in total of (5). While, for the direct optimization of the combination using (6), we have to compare all the possible associations of corrections for sources S_1 and S_2 (CD-CD, CD-CR, CD-CN, CR-CD, CR-CR, CR-CN, CN-CD, CN-CR and CN-CN) leading then to a richer frame of possible corrections, but with more comparisons to do, $3^2 = 9$ in this scenario.

In the following section, we show with several experiments both on synthetic and real data, that this second scheme can have an interest due to its performances.

4 Experiments

To test these schemes (individual corrections - Figure 1 - vs global correction - Figure 2), several numerical experiments conducted on synthetic and real data sets using two evidential classifiers are exposed in this Section.

The first classifier is the evidential k-nearest neighbor (EkNN) [3, 6] with $k = 5$. The second chosen classifier is the evidential neural networks (ENN) [4, 6] with number of prototypes $np = 5$.

For each data set, the following experiment was repeated 10 times:

- One half of the data ($\mathcal{L}_1$) is used to learn the classifier (EkNN or ENN);
- A 10-fold cross validation is then performed on the second half of the data with 9 folds ($\mathcal{L}_2$) to learn the best correction, and 1 fold for testing.

The synthetic data set, illustrated in Figure 3, has been generated by multivariate normal distribution composed of 2 features, 900 objects and 3 classes with the means $\mu_1 = (0, 2), \mu_2 = (1, 3), \mu_3 = (2, 2)$ and the following covariance matrices for each class: $\Sigma_1 = 0.1I$, $\Sigma_2 = 0.5I$ and $\Sigma_3 = \begin{bmatrix} 0.3 & -0.15 \\ -0.15 & 0.3 \end{bmatrix}$, where I is the 2×2 identity matrix.

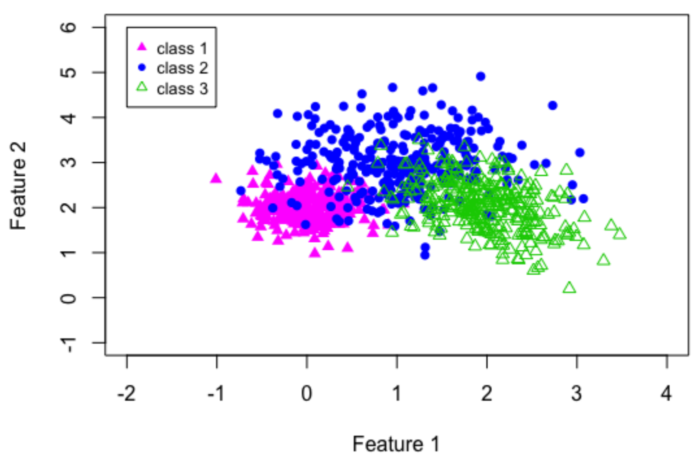

Fig. 3. Generated data set.

The real data sets used were taken from UCI [8]. Their descriptions can be seen in Table 2.

The results are summed up in Table 3 using as a measure of performance E_{pl} (5), meaning the squared error between the contour function resulting from the combination and the indicator function of the truths of the objects in the test set.

It can be seen from Table 3 that the second scheme optimizing the combination reaches better performances according to E_{pl} (5) than the first scheme combining individual optimizations.

Table 2. Description of the UCI data sets used [8]

Data sets	#Instances	#Features	#Classes
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

Table 3. Performances (Average E_{pl} (5) values), the lower the better, obtained from two sources (EkNN and ENN) for a conjunctive combination without correction (No correction), for scheme 1 (best individual corrections), for scheme 2 (best parameterized combination), for scheme 2 with only CD, only CR and only CN to highlight the interest of possibly using multiple distinct corrections. Standard deviations are indicated in parentheses. In bold the best performance for each data set.

Data	No correction CC	Scheme 1	Scheme 2	Scheme 2 only CD	Scheme 2 only CR	Scheme 2 only CN
Synthetic	10.953 (3.094)	11.690 (2.748)	9.496 (2.492)	**9.491 (2.525)**	10.961 (3.076)	10.882 (2.913)
Haberman	6.054 (2.381)	6.742 (1.664)	**5.515 (1.725)**	5.886 (2.096)	5.717 (2.070)	5.577 (1.849)
Iris	0.503 (0.619)	0.569 (0.608)	**0.467 (0.715)**	0.471 (0.712)	0.503 (0.619)	0.503 (0.619)
Glass	5.016 (1.765)	6.007 (0.778)	**4.703 (1.340)**	4.965 (1.737)	4.771 (1.408)	4.763 (1.267)
Ionosphere	2.411 (0.882)	2.874 (0.831)	**2.057 (0.958)**	**2.057 (0.958)**	2.411 (0.882)	2.411 (0.882)
Lympho	2.305 (1.077)	2.748 (0.929)	2.253 (1.058)	**2.239 (1.045)**	2.322 (1.086)	2.322 (1.059)
Liver	7.937 (1.778)	9.481 (0.942)	**7.515 (1.242)**	7.728 (1.564)	7.848 (1.565)	7.743 (1.405)
Pima	13.123 (2.770)	16.408 (1.970)	**12.455 (2.415)**	12.489 (2.490)	13.130 (2.738)	13.107 (2.681)
Sonar	3.489 (1.071)	4.365 (0.834)	3.164 (0.918)	**3.160 (0.969)**	3.491 (1.037)	3.492 (0.992)
Transfusion	17.018 (3.798)	16.393 (2.240)	**13.218 (2.230)**	15.561 (3.111)	15.096 (2.944)	13.528 (2.190)
Vehicles	23.277 (3.161)	30.886 (1.220)	**21.741 (2.209)**	23.106 (3.193)	22.949 (2.687)	21.851 (2.166)
Vertebral	4.632 (1.777)	5.173 (1.600)	**3.979 (1.512)**	3.995 (1.525)	4.645 (1.763)	4.573 (1.576)

We also wanted to highlight the possible interest of taking advantage of using possibly several different corrections and so the performances of scheme 2 with only CD, only CR and only CN were also exposed for comparisons. Using only one kind of correction can limit the performances.

It can be observed that it happens that scheme 2 with only CD (Scheme 2 only CD) obtains slightly better performances (on Iris, Lympho and Sonar data) than scheme 2 testing all combinations including CD-CD. Several non-exclusive explanations may be given: first, the best configuration on the training set is not necessarily the best one on the test set; second, the optimization on the learning set is only local; and at last, the performance measure E_{pl} (5) is somewhat favorable to CD (Details in [14, Section 8.5.1]).

As expected, the drawback to reach these performances is a longer time to learn the parameters as shown in Table 4. With only two sources, this time remains reasonable. If the number of sources were to become too large, it would certainly be necessary to see if Scheme 2 is still applicable within a reasonable time.

Table 4. Time consumption in seconds on a macbook Air M1 3.2 GHz 8 GB RAM for the learning phase for Scheme 1 and Scheme 2. Standard deviations are indicated in parentheses.

Data	Scheme 1	Scheme 2
Synthetic	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)

We now give the results according to another performance measure, and to consider the interest of belief function modeling, we look at partial decisions (meaning decision possibly in favor of a group of classes) [7], and we consider that the set of possible decisions (or acts) is equal to Ω, so we can use [12][Page 6, Strong dominance criterion with $0-1$ utilities and pieces of information represented by belief functions] the following relation of dominance between the singletons of Ω:

$$\omega \succeq \omega' \iff Bel(\{\omega\}) \geq Pl(\{\omega'\}) \ , \tag{7}$$

and make a partial decision composed of the non dominated singletons according to relation (7).

The results are then exposed in Table 5 using the u_{65} utility measure. This measure, introduced by Zaffalon et al. [20], allows one to take into account the interest of partial decisions for preferring the imprecision to being randomly correct.

The U_{65} value of a partial decision d, possibly in favor a set of singletons, is formally defined by

$$U_{65}(x) = 1.6x - 0.6x^2 \tag{8}$$

with x the so called discounted accuracy of d defined by $\frac{\mathbb{I}(\omega \in d)}{|d|}$, with $\mathbb{I}$ the indicator function, ω the true class of the instance, and $|d|$ the number of elements in d. The u_{65} utility measure gives a greater utility to imprecise but correct partial decisions of size n (meaning decisions equal to a set of n singletons one of them being the true class) than precise decisions (in favor of one singleton) only randomly correct with probability $\frac{1}{n}$.

Table 5. Performances (Average U_{65} values), the higher the better, obtained from two sources (EkNN and ENN) for a conjunctive combination without correction (No correction), for scheme 1 (best individual corrections), for scheme 2 (best parameterized combination), for scheme 2 with only CD, only CR and only CN to highlight the interest of possibly using multiple distinct corrections. Standard deviations are indicated in parentheses. In bold the best performance for each data set.

Data	No correction	Scheme 1	Scheme 2	Scheme 2 only CD	Scheme 2 only CR	Scheme 2 only CN
Synthetic	0.850 (0.052)	0.857 (0.046)	0.857 (0.046)	**0.858 (0.046)**	0.850 (0.052)	0.850 (0.052)
Haberman	0.755 (0.103)	0.747 (0.112)	0.758 (0.096)	**0.762 (0.097)**	0.756 (0.103)	0.759 (0.104)
Iris	**0.971 (0.062)**	**0.971 (0.062)**	0.968 (0.064)	0.969 (0.063)	**0.971 (0.062)**	**0.971 (0.062)**
Glass	0.677 (0.151)	0.679 (0.148)	0.681 (0.149)	0.684 (0.144)	**0.688 (0.150)**	0.683 (0.146)
Ionosphere	**0.938 (0.047)**	0.938 (0.046)	0.933 (0.043)	0.933 (0.043)	**0.938 (0.047)**	**0.938 (0.047)**
Lympho	**0.815 (0.139)**	0.805 (0.149)	0.805 (0.143)	0.806 (0.143)	0.814 (0.139)	**0.815 (0.135)**
Liver	0.697 (0.102)	0.684 (0.110)	0.703 (0.087)	**0.704 (0.092)**	0.700 (0.104)	0.698 (0.099)
Pima	0.769 (0.067)	0.767 (0.068)	**0.777 (0.066)**	0.775 (0.066)	0.769 (0.067)	0.771 (0.067)
Sonar	0.789 (0.132)	0.783 (0.132)	**0.812 (0.096)**	0.810 (0.096)	0.788 (0.135)	0.791 (0.131)
Transfusion	0.746 (0.060)	0.756 (0.064)	**0.768 (0.056)**	0.762 (0.053)	0.752 (0.062)	0.763 (0.057)
Vehicles	0.623 (0.062)	0.614 (0.064)	**0.635 (0.058)**	0.631 (0.059)	0.621 (0.063)	0.633 (0.058)
Vertebral	0.809 (0.106)	0.808 (0.104)	**0.833 (0.094)**	0.833 (0.093)	0.809 (0.106)	0.816 (0.105)

At last, with a classical error rate with decisions made for example by choosing the class maximizing the pignistic probability [17], we obtain results very similar with or without correction (Due to the page limit, it is difficult to put all results). The results on this point are preliminary, we will explore other experiments with surely more classifiers to see a possible interest of this approach.

5 Conclusion

In this paper, we have illustrated through experiments the interests of using different contextual corrections to optimize the conjunctive combination of the outputs of a group of evidential sources. We have also given elements of the possible limitations of this strategy when the number of sources to be combined increases, and according to classical error rate, these limitations remaining to be more clarified. A perspective of interest and topicality will be to study the possibility of using these schemes in an end-to-end learning of a group of deep evidential classifiers in the lines of the works of Tong et al. for example [19].

Acknowledgements The authors would like to thank the anonymous reviewers for their very helpful and relevant comments.

References

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):11071136, 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. T. Denœux. Decision-making with belief functions: A review. Int. J. of Approx. Reason., 109:87–110, 2019.
8. D. Dua, C. Graff. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, 2019.
9. 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.
10. 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.
11. L.I. Kuncheva. Combining Pattern Classifiers: Methods and Algorithms, 2nd Edition. Wiley, Hoboken, New Jersey, United States, 2014.
12. L. Ma, T. Denœux. Partial classification in the belief function framework. Knowledge-Based Syst., 214: article 106742, 2021.
13. 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.
14. 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.
15. P. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason., 9(1) : 1-35, 1993.
16. P. Smets. Analyzing the combination of conflicting belief functions, Inf. Fusion, 8(4):387–412, 2007.
17. P. Smets, R. Kennes. The Transferable Belief Model. Artificial Intelligence., 66(2):191-234, 1994.
18. G. Shafer. A mathematical theory of evidence. Princeton University Press, Princeton, N.J, 1976.
19. Z. Tong, P. Xu, T. Denœux. Fusion of Evidential CNN Classifiers for Image Classification. In T. Denœux, E. Lefvre, Z. Liu and F. Pichon (Eds), Belief Functions: Theory and Applications, Springer International Publishing, pages 168-176, 2021.
20. M. Zafallon, G. Corani, D.-D. Mauá. Evaluating credal classifiers by utility-discounted predictive accuracy. Int. J. Approx. Reason., 53(8):1282-1301, 2012.