Improving an Evidential Source of Information Using Contextual Corrections Depending on Partial Decisions

Siti Mutmainah, Samir Hachour, Frédéric Pichon, David Mercier

Improving an evidential source of information using contextual corrections depending on partial decisions[*]

Siti Mutmainah[1,2], Samir Hachour[1], Frédéric Pichon[1], and David Mercier[1]

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

Abstract. In this paper, an improvement of the quality of an evidential source of information is proposed using contextual corrections depending on partial decisions obtained from an interval dominance relation on the source outputs. Numerical experiments with the EkNN classifier and synthetic and real data allows us to illustrate the performances and the interest of this method.

Keywords: Belief functions · Contextual corrections · Partial decisions · Interval dominance

1   Introduction

In pattern recognition [1, 9], the quality of the information provided by a source (*e.g.* a sensor, a classifier, . . .) plays an important role in the success of the pattern recognition task as the information may be false, biased or irrelevant.

The belief function framework (or Dempster-Shafer theory [18]) provides a flexible mathematical framework for dealing with imperfect information. In this theory, the quality of a source of information is classically managed by means of the discounting operation introduced by G. Shafer in his seminal book [18, chapter 11, page 251]. This method has since been refined using so-called contextual correction mechanisms [12, 16] taking into account more refined knowledge about the quality of a source: its reliability (in the sense of its relevance, meaning the capacity of the source to answer the question of interest) and its truthfulness (meaning its capacity to tell what it knows; this capacity being possibly conscious - as a lie for example - or unconscious - as a bias for example) [15]. More specifically, three mechanisms have been introduced. They are respectively called *Contextual Discounting (CD)*, *Contextual Reinforcement (CR)* and *Contextual Negating (CN)*. They all can be mathematically derived [16] from these notions of reliability and truthfulness: CD, which generalizes the discounting operation,

can adjust the output of a source in accordance with information about its reliability, while CN, which generalizes the negation of a source [8], can adjust the source according to its truthfulness, and at last, CR is the dual operation of CD, it may reinforce a too cautious source [14, 16].

In this paper, we propose to improve the quality of a source of information outputting belief functions regarding a question of interest using contextual corrections CD, CR or CN, and partial decisions computed from the outputs using the relation of interval dominance [19, 20], also called strong dominance in [6, 11]. More specifically, the source is considered as a black box meaning we have no access to the manner it works to make its evidential outputs. This situation occurs for example when a company buys a sensor from another one to perform a given task, and the decision making process (or the algorithm) used by this sensor is protected [13]. A learning set is available. It is composed of outputs of the source (as mass functions for example), regarding data the ground truth of which is known. As a simple example [10], we may have in the learning set the following information m_S output by the source regarding the true class of an object o, which belongs to a universe $\Omega = \{\omega_1, \omega_2, \omega_3\}$

	ω_1	ω_2	ω_3	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_3\}$	$\{\omega_2, \omega_3\}$	Ω	Ground truth of object o
$m_S\{o\}$	0	0	0.5	0	0	0.3	0.2	ω_1

Now, instead of learning a best possible correction among CD, CR and CN from the whole learning set as proposed in [16], we would like to propose a method that takes advantage of all the corrections, and so propose to regroup the outputs leading to the same partial decisions to learn the best possible correction between CD, CR and CN in each of these groups of outputs to reach better global performances (certainly at the cost of learning more models). With this strategy, an output is thus adjusted differently depending on the partial decision it leads to.

This paper is organized as follows. In Section 2, the basic concepts and notations on belief functions used in this paper are presented, as well as a reminder on decision making with interval dominance with belief functions. Reminders on contextual corrections are given in Section 3. Thereafter, in Section 4, the proposed method to learn contextual corrections depending on partial decisions is exposed. It is tested with synthetic and real data in Section 5. A discussion is also added in this last Section to conclude the paper.

2   Belief functions: necessary concepts and notations

In this Section, necessary concepts and notations used in this paper are quickly reminded. Further details can be found for example in [18, 17, 4].

With $\Omega = \{\omega_1, ..., \omega_K\}$ the universe (or the frame of discernment) representing the finite set of answers to a given question of interest, a piece of evidence regarding the answer to this question of interest induces a *mass function (MF)* m^Ω (or m if no ambiguity) defined from 2^Ω to $[0, 1]$, verifying $\sum_{A \subseteq \Omega} m^\Omega(A) = 1$.

The *focal elements* of a MF m are the subsets $A \subseteq \Omega$ s.t. $m(A) > 0$. A MF having only one focal element A is called a *categorical MF* and can be simply denoted by m_A.

A MF m is in one-to-one correspondence with *belief and plausibility functions* Bel and Pl respectively defined for all $A \subseteq \Omega$ by $Bel(A) = \sum_{B \subseteq A} m(B)$, and $Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$.

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

If a source S provides an output m_S, and if it is also known that this source is reliable with a degree of belief $\beta = 1 - \alpha \in [0, 1]$, then this original MF m_S can be *discounted* into a MF m s.t.

$$m = \begin{cases} A \mapsto \beta m_S(A) & \forall A \subset \Omega \\ \Omega \mapsto \beta m_S(\Omega) + \alpha \end{cases} \tag{1}$$

Equation (1) can also be simply rewritten as $m = \beta\, m_S + \alpha\, m_\Omega$, with m_Ω the categorical MF defined by $m_\Omega(\Omega) = 1$, and it can also be shown (see for example [16, Prop. 11]) that the contour function associated with the discounted MF is defined for all $\omega \in \Omega$ by $pl(\omega) = 1 - (1 - pl_S(\omega))\beta$, with pl_S the contour function of m_S. Derivations of this operation can be found in [17, 12, 16].

At last, when a decision has to be made[3, 11], if we consider that the set of possible decisions (or acts) is equal to Ω, we can use the following relation of dominance between the singletons of Ω:

$$\omega \succeq \omega' \iff Bel(\{\omega\}) \geq Pl(\{\omega'\}) , \tag{2}$$

and make a partial decision composed of the non dominated singletons according to relation (2).

Due to lack of space, details cannot be written, but Equation (2) comes from for example [6][Equation 43] or [11][Page 6, Strong dominance criterion] with $0 - 1$ utilities and pieces of information represented by belief functions.

3   Contextual corrections and learning

In this section, the definitions of CD, CR and CN are recalled as well as the possibilities to learn them from labelled data [16] composed of the outputs of a source regarding objects whose true class is known.

The contour functions resulting from CD, CR and CN are recalled with K parameters, K being the number of elements in Ω. These mechanisms can indeed be used with more parameters [12, 16] but as shown in [16], these configurations with K parameters for each corrections are expressive enough to reach the lowest possible values (with these mechanisms) of the following measure of discrepancy [12, 16] between the source outputs adjusted with these corrections and the ground truth:

$$E_{pl}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} (pl_i(\omega_k) - \delta_{i,k})^2 , \tag{3}$$

where n is the number of objects in the learning set, $\boldsymbol{\beta} = (\beta_\omega, \omega \in \Omega)$ is the vector composed of the K parameters of each correction, pl_i is the contour function regarding the class of the object i resulting from a contextual correction (CD, CR or CN) of the MF provided by the source for this object, and $\delta_{i,k}$ is the indicator function of the truth of all the instances $i \in \{1, \ldots, n\}$, meaning $\delta_{i,k} = 1$ if the class of the instance i is ω_k, otherwise $\delta_{i,k} = 0$.

The discrepancy measure E_{pl} yields a linear least-squares optimization problem, which can be then efficiently solved using standard algorithms.

As for the discounting (1), we consider a source of information outputting a MF m_S regarding a question of interest. The corresponding contour functions of each contextual correction of m_S are summed up in Table 1.

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

Corrections	Contour functions
CD	$pl(\omega) = 1 - (1 - pl_S(\omega))\beta_\omega$
CR	$pl(\omega) = pl_S(\omega)\beta_\omega$
CN	$pl(\omega) = 0.5 + (pl_S(\omega) - 0.5)(2\beta_\omega - 1)$

4   Contextual corrections depending on partial decisions

In this paper, we propose to improve the previous learning method exposed in Section 3, meaning we would like to reach better performances. For this, we will consider, in the learning set, groups of distinct partial decisions the outputs of the source lead to.

The idea is to consider that the quality of the source, and then the way we have to adjust it, may depend on the outputs it gives regarding the objects whose true classes are to be found. For example, the source may be quite right when it decides a certain type of class, while having some bias when it decides another class or a group of classes, and another bias for another class or group of classes, etc.

Specifically we investigate the quality of the source depending on groups of partial decisions according to relation (2). CD, CR and CN best parameters $\boldsymbol{\beta}$ according to E_{pl} (3) are then computed in each group. The correction with the lowest value of E_{pl} is kept in each group.

This learning procedure is summarized by Algorithm 1.

5   Numerical Experiments and Discussion

In this Section we present experiments made with the EkNN classifier [2, 5] with $k = 5$ as the source of information on synthetic and real data.

Algorithm 1 Learning procedure

Input: A set $\mathcal{I}$ of instances $(m_i\{o_i\}, \omega_i)$, $i \in \{1, \ldots, n\}$, $m_i\{o_i\}$ being the output of the source regarding object o_i whose true class is ω_i.
Outputs: Groups $\mathcal{G}$ of partial decisions and best corrections for each group.

1: **procedure** LEARNINGPROCEDURE
2:     $\mathcal{G}$ initially empty
3:     **for** each instance i in $\mathcal{I}$ **do**
4:         Compute the partial decision coming from m_i using relation (2)
5:         Add i to the group g in $\mathcal{G}$ associated with this partial decision.
6:     **for** each group of partial decisions g in $\mathcal{G}$ **do**
7:         Compute CD, CR and CN best parameters $\boldsymbol{\beta}$ according to E_{pl} (3) restricted to the instances in this group
8:         Keep for this group the correction reaching the lowest value of E_{pl}.

Synthetic data are composed of 5000 instances, 5 classes and 2 features, which were generated from a multivariate normal distribution with means $\mu_1 = (0,0), \mu_2 = (2,0), \mu_3 = (0.2), \mu_4 = (2,2), \mu_5 = (1,1)$ for respectively class 1, 2, 3, 4 and 5, and the same covariance matrix Σ for each class:

$$\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \tag{4}$$

An illustration of a generated synthetic data set with these parameters is given in Figure 1.

The real data sets used from UCI [7] are described in Table 2.

Table 2: Descriptions of used UCI data sets.

Data sets	# instances	# features	# classes
Breast cancer	569	31	2
Glass	214	10	6
Haberman	306	3	2
Ionosphere	350	34	2
Iris	150	4	3
Liver	345	6	2
Lymphography	146	18	3
Pima	768	8	2
Red wine	1599	12	6
Sonar	208	60	2
Transfusion	748	3	2
Vehicles	846	19	4
Vertebral	310	6	3

For each data set, the following experiment is repeated 10 times: one half of the data ($\mathcal{L}_1$) is used to learn the EkNN classifier; then a 10-fold cross validation
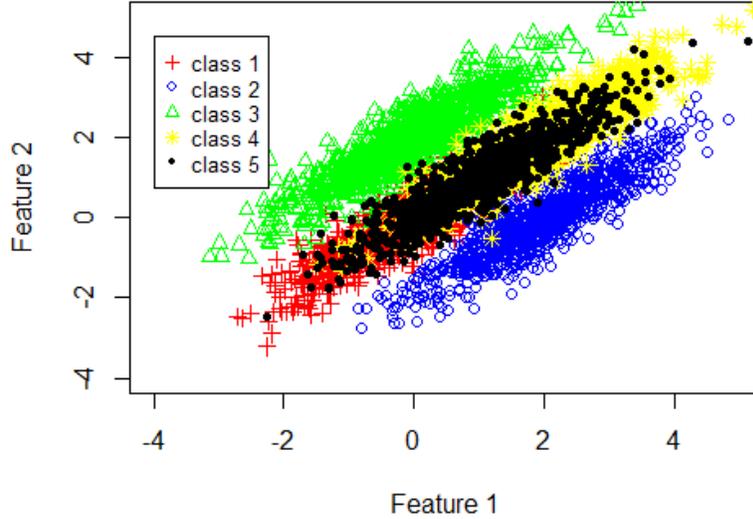
Fig. 1: Example of a generated data set.

is performed on the second half of the data with 9 folds ($\mathcal{L}_2$) to learn the best correction in each group of partial decisions using Algorithm 1, and 1 fold for testing (meaning for the test phase).

Some possible partial decisions obtained from the outputs of the source (the EkNN classifier) with the generated data set illustrated in Figure 1 are given in Figure 2 for some points in the feature space. Note that these groups of partial decisions are not necessary computed in the training phase $\mathcal{L}_2$. They are just given here as an illustration. As exposed in Algorithm 1, only the points (and their associated partial decisions) in the fold $\mathcal{L}_2$ are considered to learn the corrections. That is why it may happen that during the test phase, a partial decision did not happen in the learning phase. In this situation, we propose to use the best correction for the whole set of $\mathcal{L}_2$. Note that in our experiments presented here, this case very rarely occurred.

To measure the performances, we used the measure E_{pl} (3) with which corrections are learnt, and we also wanted to use another measure with which corrections are not learnt. For this second choice, we opted for the u_{65} utility measure, introduced by Zaffalon et al. [21], which allows one to take into account the advantages of partial decisions concerning the fact of preferring imprecision to being randomly correct. The u_{65} utility measure gives indeed a greater utility to imprecise but correct partial decisions of size n (meaning decisions equal to
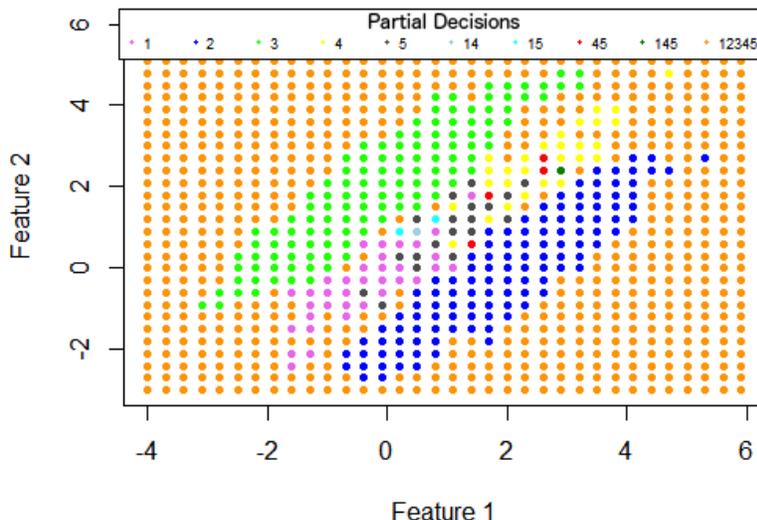
Fig. 2: Points in the feature space belonging to different groups of partial decisions obtained from the generated synthetic data set composed of 5 classes and 2 attributes/features illustrated in Figure 1. The legend for the partial decisions obtained from the EkNN classifier and relation (2) are given above the figure. We can see that for these points, 10 possible partial decisions have been obtained: 1 (meaning a decision for class 1), 2, 3, 4, 5, 14 (meaning a partial decision in favor of class 1 or class 4), 15, 45, 145 (meaning a partial decision in favor of class 1 or class 4 or class 5), 12345 (meaning a partial decision in favor of all the classes - total uncertainty - each class can be the good one).

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 $\frac{1}{n}$.

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

$$U_{65}(x) = 1.6x - 0.6x^2 \tag{5}$$

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.

Performances according to E_{pl} (3) and the averages of the u_{65} utility measures [21] of the partial decisions obtained from the corrected outputs are regrouped in Table 3.

Table 3: Performances (Average E_{pl} values, the lower the better, and average U_{65} utilities, the greater the better) obtained for EkNN learnt with $\mathcal{L}_1$ (denoted by EkNN), EkNN learnt with groups $\mathcal{L}_1$ and $\mathcal{L}_2$ (denoted by EkNN+), EkNN learnt with $\mathcal{L}_1$ followed by corrections (CD, CR, CN) learnt with $\mathcal{L}_2$, and EkNN learnt with $\mathcal{L}_1$ followed by the new method using groups of partial decisions learnt with $\mathcal{L}_2$. Standard deviations are indicated in parenthesis.

Data	EkNN	EkNN+	CD	CR	CN	Method
Average E_{pl} values (the lower the better)						
Synthetic	188.27 (26.48)	181.27 (16.23)	187.91 (26.56)	134.57 (9.83)	174.06 (21.38)	**106.76 (7.29)**
Breast cancer	3.74 (2.50)	3.59 (2.43)	3.65 (2.35)	3.64 (2.34)	3.56 (2.22)	**3.56 2.21)**
Glass	8.16 (3.61)	8.97 (3.19)	8.13 (3.57)	5.11 (1.40)	6.93 (2.04)	**4.81 (1.49)**
Haberman	8.59 (2.40)	8.79 (2.34)	8.57 (2.32)	5.97 (1.84)	6.61 (1.07)	**5.87 (1.83)**
Ionosphere	2.31 (1.66)	**1.64 (1.18)**	2.31 (1.62)	2.19 (1.49)	2.25 (1.52)	1.81 (1.49)
Iris	0.56 (0.87)	0.59 (1.02)	0.56 (0.85)	**0.55 (0.82)**	0.56 (0.84)	0.56 (0.86)
Lymphography	3.07 (1.44)	**2.32 (1.27)**	3.08 (1.44)	2.74 (1.24)	3.02 (1.18)	2.74 (1.40)
Liver	12.66 (2.08)	12.21 (1.56)	12.66 (2.02)	8.17 (1.08)	8.31 (0.64)	**8.01 (1.12)**
Pima	20.01 (3.32)	18.12 (3.21)	20.01 (3.32)	15.67 (2.20)	16.76 (1.57)	**14.93 (2.40)**
Red wine	162.09 (21.75)	140.60 (13.98)	162.09 (21.75)	48.62 (2.96)	116.83 (4.28)	**45.92 (3.44)**
Sonar	4.04 (1.43)	**3.06 (1.33)**	4.05 (1.43)	3.50 (1.08)	3.80 (1.05)	3.30 (1.37)
Transfusion	20.54 (4.91)	19.77 (4.37)	19.43 (4.19)	15.21 (3.45)	15.44 (2.21)	**13.56 (2.85)**
Vehicles	34.26 (5.66)	28.14 (4.74)	34.26 (5.66)	24.07 (2.35)	32.40 (4.20)	**20.69 (2.96)**
Vertebral	5.08 (2.29)	4.68 (2.05)	5.04 (2.19)	4.64 (1.89)	4.85 (1.89)	**4.31 (1.81)**
Average U_{65} values (the greater the better)						
Synthetic	66.52 (2.68)	66.40 (2.63)	66.49 (2.73)	66.10 (2.69)	65.46 (2.89)	**68.36 (2.59)**
Breast cancer	91.77 (4.65)	**93.74 (3.90)**	92.76 (4.60)	92.75 (4.64)	92.80 (4.62)	92.76 (4.60)
Glass	61.51 (14.47)	63.38 (13.44)	61.56 (14.25)	64.53 (13.87)	45.93 (13.22)	**66.24 (12.35)**
Haberman	74.28 (8.12)	74.45 (8.57)	74.61 (7.97)	74.84 (9.80)	71.69 (8.04)	**75.13 (9.26)**
Ionosphere	93.02 (4.67)	**94.71 (4.16)**	93.02 (4.67)	93.12 (4.62)	92.98 (4.64)	93.86 (4.99)
Iris	96.78 (6.78)	96.51 (6.07)	96.78 (6.78)	96.84 (6.72)	96.78 (6.78)	**97.06 (6.20)**
Liver	66.74 (5.86)	**67.14 (5.53)**	66.73 (5.86)	66.64 (6.82)	64.70 (2.25)	65.45 (7.42)
Lymphography	80.23 (13.13)	**83.84 (12.65)**	80.21 (13.20)	79.42 (13.65)	79.26 (12.26)	79.11 (14.49)
Pima	72.64 (5.06)	73.00 (5.56)	72.64 (5.06)	73.34 (5.50)	71.97 (3.91)	**73.38 (5.02)**
Red wine	45.39 (4.17)	53.29 (4.23)	45.39 (4.17)	57.60 (4.14)	25.00 (0)	**59.12 (4.12)**
Sonar	78.34 (8.53)	**84.50 (7.69)**	78.27 (8.60)	78.96 (8.95)	78.00 (8.03)	79.32 (10.49)
Transfusion	72.76 (5.68)	72.87 (5.41)	74.20 (5.29)	74.39 (6.38)	72.63 (5.42)	**75.91 (6.03)**
Vehicles	61.20 (5.77)	63.46 (5.62)	61.20 (5.77)	60.56 (5.65)	58.11 (5.60)	**63.61 (6.20)**
Vertebral	80.40 (10.14)	**82.09 (8.85)**	80.29 (10.04)	79.95 (10.27)	80.36 (10.19)	81.78 (9.40)

As it can be seen in Table 3, the proposed method using the groups of partial decisions obtains almost in each situation better results than the previous learning using only CD or CR or CN.

Furthermore, the source was supposed to be a black box, but we were also curious to see the performances of this source if we were able to improve its performances by using the data we use to learn the corrections. These performances are given in the column $EkNN+$ in Table 3. We can see that the new learning method succeeds to even improve these results for some data sets.

As summarized in [11] by Ma and Denœux, relation (2) is only one among others. First future works will then consist in testing these other possible relations (s.t. weak dominance, maximality, ...) to compute the groups of partial decisions and see if better performances can be reached.

References

1. C.M. Bishop. Pattern Recognition and Machine Learning. Springer, New York, United States, 2006.
2. T. Denœux. A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Trans. Syst. Man Cybern., 25(5):804-813, 1995.
3. T. Denœux. Analysis of evidence-theoretic decision rules for pattern classification. Pattern recognit., 30(7):1095–1107, 1997.
4. T. Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artif. Intell., 172:234-264, 2008.
5. T. Denœux. Evclass: Evidential distance-based classification, [https://cran.r-project.org/web/packages/evclass/index.html], R package version 1.1.1. 2017.
6. T. Denœux. Decision-making with belief functions: A review. Int. J. of Approx. Reason., 109:87–110, 2019.
7. D. Dua, C. Graff. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, 2019.
8. D. Dubois, H. Prade. A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. Int. J. Gen. Syst., 12(3):193-226, 1986.
9. R.O. Duda, P.E. Hart and D.G. Stork. Pattern Classification. Wiley, New York, United States, 2nd Edition, 2001.
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. Ma, T. Denœux. Partial classification in the belief function framework. Knowledge-Based Syst., 214: article 106742, 2021.
12. 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.
13. D. Mercier, G. Cron, T. Denux, M.-H. Masson. Decision fusion for postal address recognition using belief functions. Expert Syst. Appl., 36(3):5643–5653, 2009.
14. D. Mercier, E. Lefèvre, F. Delmotte. Belief functions contextual discounting and canonical decompositions. Int. J. Approx. Reason., 53(2):146-158, 2012.
15. F. Pichon, D. Dubois, T. Denoeux. Relevance and truthfulness in information correction and fusion. Int. J. Approx. Reason., 53(2):159-175, 2012.
16. 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.
17. P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason., 9(1) : 1-35, 1993.
18. G. Shafer. A mathematical theory of evidence. Princeton University Press, Princeton, N.J, 1976.
19. M.C. Troffaes. Decision making under uncertainty using imprecise probabilities. Int. J. Approx. Reason., 45(1):17–29, 2007.
20. G. Yang, S. Destercke, M.H. Masson. Nested Dichotomies with probability sets for multi-class classification. ECAI, pages 363–368, 2014.

21. M. Zafallon, G. Corani, D.-D. Mauá. Evaluating credal classifiers by utility-discounted predictive accuracy. Int. J. Approx. Reason., 53(8):1282-1301, 2012.