Interval Criterion-Based Evidential Set-Valued Classification

Abdelhak Imoussaten[1][0000−0002−1292−2681] and Jacky Montmain[1][0000−0003−0918−5788]

EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France
{abdelhak.imoussaten,jacky.montmain}@mines-ales.fr

Abstract. This paper deals with set-valued classification methods. The aim of these methods is to provide a subset of classes as a prediction that is cautious but not too large. The well known *Strong Dominance* based set-valued classification algorithm (SD) is a good candidate as a robust method but sometimes the predicted subsets are too large. This paper proposes a flexible method that is a trade-off between SD based method and a point classification method. Indeed, the proposed set-valued classifier within the framework of belief functions, called IC, controls the granularity of the partial order by predicting a compromise between the cautiousness offered by the SD and the precision offered by point prediction classifiers. It is based on a interval criterion that is built from the pignistic criterion to which is associated a threshold. The introduced threshold aims to incorporate the decision-maker preference regarding the data imperfections. The paper shows the management of the interval comparisons and the intransitive binary relations resulting from the introduction of the threshold using graph theory and decision theory. The outputs of the IC are theoretically studied and compared to the prediction of SD and the pignistic criterion. Therefore, its performances regarding five set-valued classification performances measures are compared using fashion mnist image data. Experimental results show that IC gives good performances following trade-off measures.

Keywords: Set-valued classification · Interval Criterion · Belief Functions Theory.

1 Introduction

Set-valued prediction methods, also called imprecise or non-deterministic prediction methods, refers to the classification methods that return a subset of classes/interval instead of betting on a single class/value in presence of uncertainty and/or imprecision. These methods are necessary when a classification task is involved in applications that are sensitive and where each error has serious consequences, as in medical diagnosis applications; or in autonomous car applications; or in environmental compliance [9], etc. The challenge of these methods is to return a subset of classes that contains the true class while remaining as small as possible. For this aim, uncertainty and imprecision should be quantified in the trained model and in the object to classify. Recently several propositions were made within different uncertainty framework. Within the framework of probability theory, one can cite the *non-deterministic* classifier (ndc) [4] or by using statistical hypothesis testing to provide confidence regions as conformal prediction [15], within the framework of imprecise probabilities, one can find the classifiers *Naive Credal Classifier* (*ncc*) [18], the credal decision trees (CDT) [2] and (ICDT) [1]. Within the framework of belief function it exists two main categories which are *weak order* and *partial preorder* based approaches where criteria like the *generalized maximin criterion* and *generalized maximax criterion* are used in the first approach and criterion like *strong dominance* is used in the second approach [5] [14] [11] [7]. In this paper we focus, within the framework of belief function, on set-valued classifiers that provide a prediction for a sample $\boldsymbol{x}$ based on the mass function, denoted $m_{\boldsymbol{x}}$, quantifying the chances of each subset, from the set of classes, to contain the true class of $\boldsymbol{x}$. More precisely, we are interested on the *strong dominance* based classifier (SD) which gives very good performances concerning the cautiousness measure. The idea of using SD binary relation is justified by the fact that we suspect imprecision and/or uncertainty in the data and we seek for a prediction that is robust enough. However, this robustness is at the expense of the precision, i.e., the size of the predicted subsets are, some times, too large and needs to be improved. The proposition of this paper, based on interval criterion around the pignistic criterion, aims to consider an extension of the SD binary relation that is more flexible and allow to control the granularity of the non-dominated classes returned by the SD based classifier. As the set-valued classification corresponds to the problem of selecting a subset from a large set of classes, this problem is widely studied in graph theory and decision theory [6] as the *selection problem*, we relied on these areas to make our propositions.

2 Theoretical Background

2.1 Selection Problem

2.1.1 Acts and binary relations In the context of uncertainty, the alternatives from which the decision-maker has to select the one (or ones) that satisfies his preferences are formalized by acts. An act f is an application that associates for each relevant event a consequence. Formally, let us consider a reference set Θ representing all the *states of the world* and a set of consequences $\mathcal{C}$.

Definition 1. *Let $\mathbb{V}$ be a finite set of acts. A binary relation $\mathcal{R}$ on the set $\mathbb{V}$ is a subset of the Cartesian product $\mathbb{V} \times \mathbb{V}$, that is, a set of ordered pairs (a, b) such that $a, b \in \mathbb{V}$.*

The binary relation $\mathcal{R}$ is called: *reflexive* if $\forall a \in \mathbb{V}$, $a\mathcal{R}a$; *complete* if $\forall a, b \in \mathbb{V}$, $a \neq b$, ($a\mathcal{R}b$ or $b\mathcal{R}a$); *antisymmetric* if $\forall a, b \in \mathbb{V}$, if $a\mathcal{R}b$ and $b\mathcal{R}a$ then $a = b$; *transitive* if $\forall a, b, c \in \mathbb{V}$, $a\mathcal{R}b$ and $b\mathcal{R}c$ then $a\mathcal{R}c$; *Ferrers* if $\forall a, b, c \in \mathbb{V}$, ($a\mathcal{R}b$ and $c\mathcal{R}d$) then ($a\mathcal{R}d$ or $c\mathcal{R}b$).

Definition 2. *A pair of binary relations* (I, P) *over* $\mathbb{V}$ *is said to be a* ***total preorder*** *(also called* ***weak order****) if the binary relation* $S = I \cup P$ *is reflexive, complete and transitive.*

In some situations, the presence of imprecision and uncertainty could lead the decision-maker to have imprecise preference or hesitation between two acts and therefore other concepts as interval criterion, pre-criterion and pseudo-criterion are more relevant [6]. Consequently, the incomparabilities could exist when comparing act [13] as the the binary relation is not complete. The decision-maker must then content himself with a selection of a subset of acts which contains the best solutions (*partial order*).

Definition 3. *A pair of binary relations* (I, P) *over* $\mathbb{V}$ *is said to be a* ***partial order*** *if the binary relation* $S = I \cup P$ *is reflexive, transitive, antisymmetric and not complete.*

When modelling preferences, P corresponds to the binary relation representing "strict preference" and I to the binary relation representing "indifference" and they have to be transitive in order to correspond to a *weak order* or *partial preorder*. But to cover a large real-world situations, transitivity is not always required for the binary relation I [13]. This remark is at the origin of introducing the following interval order.

Definition 4. *A pair of binary relations* (I, P) *over* $\mathbb{V}$ *is said to be an* ***interval order*** *if the binary relation* $S = I \cup P$ *is reflexive, complete and Ferrers.*

The order defined in Definition 4 admits that the corresponding symmetric binary relation I is intransitive because of the introduction of thresholds on the decision criteria. The problem that arises then is how to determine the subset of the "best" acts when such binary relations are considered? The answer to this question is to determine the kernel of the directed graph that represents the binary relation S [12].

2.1.2 The kernel of a directed graph

Let us consider a finite set of acts $\mathbb{V}$ and a binary relation S over $\mathbb{V}$. The graph associated to S is the directed graph, also called digraph, $G = (\mathbb{V}, S)$ where the **vertices**, also called nodes, are the acts of $\mathbb{V}$ and the **edges**, also called directed edges, directed links, directed lines, arrows or arcs, are the elements of S. Note that the trivial elements $(a, a) \in S$, $a \in \mathbb{V}$, are not considered when defining G. Before going any further, we will give a few reminders of some basic elements of graph theory that concern directed graphs.

Definition 5. *Let us consider a directed graph* $G = (\mathbb{V}, S)$.

- *A* ***path*** *in* G *is a succession of arcs that allows to move from one vertex to another.* G *is* ***strongly connected*** *if there is a path from* a *toward* b *and from* b *toward* a *for every* $(a, b) \in S$.

- *In G that may not itself be strongly connected, a pair of vertices a and b are said to be* ***strongly connected to each other*** *if there is a path in each direction between them. A* ***strongly connected component*** *(scc) of G is a subgraph that is strongly connected, and is maximal with this property: no additional edges or vertices from G can be included in the subgraph without breaking its property of being strongly connected.*
- *The collection of strongly connected components forms a partition of the set of vertices of G. A scc C is called* ***trivial*** *when C consists of a single vertex.*
- *G is* ***acyclic*** *if and only if all its scc are trivial.*
- *A* ***reduced graph*** $G_r = (C_r, S_r)$ *of* $G = (\mathbb{V}, S)$ *is defined as follows:*
  - $C_r = \{C_1, \ldots, C_p\}$ *where* C_i*,* $i = 1, \ldots, p$*, are the scc of G,*
  - $S_r = \{(C_i, C_j) \ / \ i \neq j, \ \exists a \in C_i \text{ and } b \in C_j \ : \ (a, b) \in S\}$.

  The directed graph $G_r = (\mathbb{V}_r, S_r)$ *is then acyclic, i.e., is formed by contracting each strongly connected component of G into a single vertex.*

Note that it is possible to test the strong connectivity of a graph, or to find its strongly connected components, in linear time, i.e., $O(|\mathbb{V}| + |S|))$.

Definition 6. *Let us consider a directed graph* $G = (\mathbb{V}, S)$*. A* ***kernel*** *of G is a subset* $\mathcal{N} \subseteq \mathbb{V}$ *such that:*

1. $\mathcal{N}$ *is absorbing: every vertex* $b \notin \mathcal{N}$ *has a predecessor in* $\mathcal{N}$*, i.e.,* $\forall b \notin \mathcal{N}$*,* $\exists a \in \mathcal{N}$ *such that* aSb.
2. $\mathcal{N}$ *is stable: N contains no pair of adjacent vertices, i.e.,* $\forall a, b \in \mathcal{N}$*,* $a \neq b$*, neither* aSb*, nor* bSa.

Note that a directed graph can have one or more kernels depending if it is acyclic or not but an acyclic directed graph has a unique kernel. So, in this case using a breadth-first type search, determining the kernel, as defined in Definition 6, requires a runtime of at worst $O(|\mathbb{V}| + |S|)$. While in the case where the graph is not acyclic, we have to determine the kernel N_r of the reduced graph G_r and if some elements of N_r belong to a non-trivial scc C, then each combination of one element from C with the others elements of $N_r \setminus C$ forms a kernel.

2.2 Belief Functions

Belief functions theory, is an interesting framework to represent and process uncertain and imprecise information. Three main set functions are involved in the belief functions framework. The *mass function* which assigns probabilities to imprecise information, leading to the distinction between equiprobability and imprecision or ignorance.

Definition 7. *The* ***mass function****, also called basic belief assignment (bba) over a finite frame of discernment* $\Theta = \{\theta_1, \ldots, \theta_n\}$*, is a set function* $m : 2^{\Theta} \to [0, 1]$ *such that* $m(\emptyset) = 0$ *and* $\sum_{A \subseteq \Theta} m(A) = 1$.

The elements $A \subseteq \Theta$ such that $m(A) > 0$, are called focal elements and they form a subset of 2^Θ denoted $\mathbb{F}$. The pair $(m, \mathbb{F})$ is called the body of evidence. The two other set functions, i.e., *belief function* and *plausibility function*, serve to make inference from the mass function m.

Definition 8. *The* ***belief function*** *quantifies the evidence proving an event,* $Bel : 2^\Theta \to [0,1]$*, satisfying for all* $A \subseteq \Theta$*,* $Bel(A) = \sum_{B \subseteq \Theta, B \subseteq A} m(B)$.

Definition 9. *The* ***plausibility function*** *quantifies the evidence that makes an event possible,* $Pl : 2^\Theta \to [0,1]$*, satisfying for all* $A \subseteq \Theta$*,* $Pl(A) = \sum_{B \subseteq \Theta, B \cap A \neq \emptyset} m(B)$.

The above-defined set functions constitute the *credal level* where beliefs are captured and quantified. A second level considered in the belief functions framework is the *pignistic level* or decision level where beliefs are quantified using probability distributions. In this second level a probability mass function is defined to make decision (or to bet).

Definition 10. *The* ***pignistic probability****, denoted* $betP_m$*, is a probability mass function defined as follows:* $\forall \theta \in \Theta$*,* $betP_m(\theta) = \sum_{A \subseteq \Theta, A \ni \theta} \frac{m(A)}{|A|}$*, where* $|A|$ *denotes the number of elements in the subset* A.

The expected utility based on the *pignistic probability* and a utility function u, gives the *pignistic criterion* (PC) $\mathbb{E}_{betP}u$ as follows:

$$\mathbb{E}_{betP}u(\theta) = \sum_{\theta' \in \Theta} betP(\theta')\ u(\theta, \theta'), \tag{1}$$

2.3 Selection Problem Within Belief Functions Framework

Let us consider a finite set of acts $\mathbb{V}$, a finite set of classes $\Theta = \{\theta_1, \ldots, \theta_n\}$ and a utility matrix u, where for $a \in \mathbb{V}$ and $\theta \in \Theta$, $u(a, \theta) \in [0,1]$ represents the utility associated to the consequence of choosing the act a if state θ occurs. Let us also consider a sample $\boldsymbol{x}$ and a mass function $m : 2^\Theta \to [0,1]$ quantifying the chances for each non-empty subset of Θ to contain the true class of $\boldsymbol{x}$. Several decision rules were proposed in the literature to decide which act to choose based on m and u. Two main approaches to select the set-valued prediction for $\boldsymbol{x}$ are to be distinguished. The first one consists in constructing a *partial preorder* over acts corresponding to elements of Θ while for the second approach the aim is to construct a *weak order* over acts corresponding to the non-empty subsets of Θ.

2.3.1 Weak-Order Based Approach In this approach, the $n \times n$ utility matrix needs to be extended to a $2^n - 1 \times n$ matrix in order to take into account the utilities $u(A, \theta)$ of predicting a non-empty subset of Θ when the true class of $\boldsymbol{x}$ is $\theta \in \Theta$. A first proposition is the one of the **ndc** classifier [4] within the Bayesian framework where the utilities $u(A, \theta)$ are computed as an extension of the F_β scores. Then two other propositions of extending u are proposed. The first one in [11] where the authors propose to generalize several classical criteria by extending u to a $2^n - 1 \times n$ matrix $\tilde{u}$ using the OWA operator [17]. The second proposition consists in a generalisation of *ndc* classifier to the case where the available information concerning the true class is considered imprecise, the **eclair GF**β set-valued classifier [7] [9] uses the extended $2^n - 1 \times 2^n - 1$ utility matrix based on a generalization of F_β scores.

2.3.2 Partial Preorder Based Approach In this approach, SD [14] is the most known set-valued classifier. It is based on the *generalized maximin criterion* (*maximin*): $\mathbb{E}u_*(\theta) = \sum_{B \subseteq \Theta} m(B) \min_{\theta \in B} u(\theta, \theta')$, and *generalized maximax criterion* (*maximax*): $\mathbb{E}u^*(\theta) = \sum_{B \subseteq \Theta} m(B) \max_{\theta \in B} u(\theta, \theta')$, and its corresponding binary relation is a partial preorder. More precisely, let us consider a $n \times n$ utility matrix u and two classes θ and θ', the binary relation denoted $\succsim_{SD}$ representing **strong dominance** is defined as follow:

$$\theta \succsim_{SD} \theta' \Leftrightarrow \mathbb{E}u_*(\{\theta\}) > \mathbb{E}u^*(\{\theta'\}). \tag{2}$$

The subset of non-dominated classes in the sense of $\succsim_{SD}$ is denoted $\mathcal{N}_{SD}$.

3 Interval-Criterion Based Set-valued Classification

3.1 Pignistic Interval-Criterion

Let us consider a set of finite classes Θ, it is also considered as the set of acts, and a mass function m defined over Θ. To compare the classes in Θ, one can use the pignistic criterion and therefore the comparisons lead to a weak order over the classes. However, pig_m is an approximation of the unknown probability mass function Pr representing the uncertainty about the classes and for each event $A \subseteq \Theta$, $Pr(A) \in [Bel_m(A), Pl_m(A)]$. In order to take into account this imprecision an interval-criterion is better suited to compare the classes. Indeed, through the concept of interval-criterion one can introduce an indifference threshold representing the biggest gap $pig_m(\theta') - pig_m(\theta)$ compatible with an indifference situation between θ' and θ. This threshold is usually used in decision theory and allows modelling the decision-maker preferences in the situations where data are subject to imprecision and uncertainties [13]. Thus, we can define the following interval-order by the pair (I_{pig}, P_{pig}) related to the pignistic interval-criterion as follows.

Definition 11. *The indifference I_{pig} and the strict preference P_{pig} binary relations are defined as follows:*

$$\begin{cases} \theta I_{pig} \theta' \Leftrightarrow \begin{cases} pig_m(\theta) - pig_m(\theta') \leq q(pig_m(\theta')) \text{ and} \\ pig_m(\theta') - pig_m(\theta) \leq q(pig_m(\theta)) \end{cases} \\ \theta P_{pig} \theta' \Leftrightarrow pig_m(\theta) - pig_m(\theta') > q(pig_m(\theta')). \end{cases} \tag{3}$$

where the threshold function q is such that $q(pig_m(\theta)) \geq 0$, $\forall \theta \in \Theta$. Note that the threshold function q depends on $pig_m(\theta)$ in the sense that when comparing act θ to the act θ', one has to take into account the imprecision about $pig_m(\theta')$ and vice versa.

3.2 Pignistic-Plausibility Interval-Criterion (2pic)

In order to make the choice of a robust method regarding this criterion, one can choose the threshold function q that ensure a strict preference for θ against θ' when $pig_m(\theta)$ is greater than the total belief that can make θ' possible, i.e., $Pl_m(\{\theta\})$. Such a proposition can be reached by considering the threshold function $q(\theta) = Pl_m(\{\theta\}) - pig_m(\theta)$. This leads to the following specific pignistic interval-criterion that is called here pignistic plausibility interval criterion (2pic):

$$\begin{cases} \theta \ I_{2pic} \ \theta' \Leftrightarrow pig_m(\theta) \leq Pl_m(\{\theta'\}) \text{ and } pig_m(\theta') \leq Pl_m(\{\theta\}) \\ \theta \ P_{2pic} \ \theta' \Leftrightarrow pig_m(\theta) > Pl_m(\{\theta'\}). \end{cases} \quad (4)$$

Proposition 1. *Let us consider the binary relation $S_{2pic} = I_{2pic} \cup P_{2pic}$ where I_{2pic} and P_{2pic} are defined as in the Equation (4). Then the following property related to the directed graph $G = (\Theta, S_{2pic})$ is verified:*

$$G \text{ is acyclic iff } I_{2pic} = \{(\theta, \theta), \theta \in \Theta\}.$$

Proof. Let us consider that G is acyclic. If $I_{2pic} \neq \{(\theta, \theta), \theta \in \Theta\}$ then it exists two different vertices $a, b \in \Theta$ such that $(a, b) \in I_{2pic}$. It comes that exists a non-trivial scc containing at least a and b. Contradiction ! Conversely, let us consider that $I_{2pic} = \{(\theta, \theta), \theta \in \Theta\}$. Suppose that it exists a non-trivial scc C in G. Then it exists two different vertices a and b of G that are in C. As, $I_{2pic} = \{(\theta, \theta), \theta \in \Theta\}$, the only two possible situations are 1) it exists $c \in C$, such that $(a, b) \in P_{2pic}$, $(b, c) \in P_{2pic}$ and $(c, a) \in P_{2pic}$ or 2) it exists $c' \in C$, such that $(b, a) \in P_{2pic}$, $(a, c') \in P_{2pic}$ and $(c', b) \in P_{2pic}$. But P_{2pic} is transitive, thus in the two situations we obtain $pig_m(a) > Pl_m(b)$ and $pig_m(b) > Pl_m(a)$. Contradiction !

Proposition 2. *Let us consider the binary relation $S_{2pic} = I_{2pic} \cup P_{2pic}$ where I_{2pic} and P_{2pic} are defined as in the Equation (4). Then the following property related to the directed graph $G = (\Theta, S_{2pic})$ is verified:*

if G contains non-trivial scc then the binary relation $S_r = I_r \cup P_r$ associated to its reduced graph $G_r = (C_r, S_r)$ (see Definition 5) is such that $I_r = \{(C_i, C_i) \ / \ C_i$ is a scc of $G\}$.

Proof. As the reduced graph $G_r = (C_r, S_r)$ of G is acyclic, it is obvious from the proposition 1 that $I_r = \{(C_i, C_i) \ / \ C_i$ is a scc of $G\}$.

Proposition 3. *Let us consider the binary relation $S_{2pic} = I_{2pic} \cup P_{2pic}$ where I_{2pic} and P_{2pic} are defined as in the Equation (4). Then the following property related to the directed graph $G = (\Theta, S_{2pic})$ is verified:*

The binary relation S_{2pic} in the case of acyclic direct graph and the binary relation $S_r = I_r \cup P_r$ associated to its reduced graph $G_r = (C_r, S_r)$ in the case of a direct graph with non-trivial scc, are total orders.

Proof. Let us consider the case when G is acyclic. Based on the proposition 1, we deduce that S_{2pic} is reflexive. S_{2pic} is as well transitive, since $S_{2pic} = P_{2pic} \cup \{(\theta, \theta), \theta \in \Theta\}$. Moreover, S_{2pic} is complete. Indeed, let us consider two different vertices a and b in Θ. Since $(a, b) \notin I_{2pic}$ then either $pig_m(a) > Pl_m(b)$, i.e., $a\ P_{2pic}\ b$, or $pig_m(b) > Pl_m(a)$, i.e., $b\ P_{2pic}\ a$. Finally, S_{2pic} is a total order. The prove still the same for the reduced directed graph in case of non-trivial scc in G.

From the Propositions 1, 2 and 3, it is obvious that if G is acyclic then the kernel of G contains exactly one vertex and if G contains non-trivial scc then the kernel of the reduced directed graph G_r contains exactly one scc C. Therefore, Each vertex of C is a kernel of G. Consequently, if G is acyclic then we can take the single class of the kernel as the output of the 2pic classifier. While in the case of the existence of non-trivial scc, if the scc containing the kernels has more than a single class, we need to decide what is the output of the 2pic classifier. In Definition 12, we suggest to select the classes of the whole scc except classes that are strictly preferred by another class in this scc.

Definition 12. *Let us consider the binary relation $S_{2pic} = I_{2pic} \cup P_{2pic}$ where I_{2pic} and P_{2pic} are defined as in the Equation 4. The resulting directed graph is denoted $G = (\Theta, S_{2pic})$.*

- *if G is acyclic then the selected class is the one in the kernel of G.*
- *if G contains non-trivial scc, let us denote $scck \subseteq \Theta$ the scc containing the kernels of G. Let us also denote by S^{scck}_{2pic} (resp. I^{scck}_{2pic} and P^{scck}_{2pic}) the restriction of S_{2pic} (resp. I_{2pic} and P_{2pic}) to $scck \times scck$. Then the subset of selected classes, denoted $\mathcal{N}_{2pic}$, is*
  - *$\mathcal{N}_{2pic} = scck$ if $P^{scck}_{2pic} = \emptyset$,*
  - *otherwise, $\mathcal{N}_{2pic} = scck \setminus \{a \in scck :\ \exists b \in scck,\ (b, a) \in P^{scck}_{2pic}\}$.*

*In the case of machine learning supervised classification, the selected subset $\mathcal{N}_{2pic}$ is called **2pic set-valued classification**.*

Example 1. Let us consider the body of evidence $(m_3, \mathbb{F}_3)$ associated to the fashion image of a Shirt that is presented in the example of Figure 2 and Table 1. Figure 1 shows the associated directed graph based on the binary relation S_{2pic} associated to the body of evidence $(m_3, \mathbb{F}_3)$. For this example, $G = (\Theta, S_{2pic})$ contains non-trivial scc and $\mathcal{N}_{2pic} = \{Dress, Shirt\}$.

3.3 Theoretical Results

Let us consider a finite set of classes Θ, and a mass function m defined over Θ. Let us denote by $\mathcal{N}_{PC}$ the subset of the best ranked classes with the *pignistic criterion*. Note that, $\mathcal{N}_{PC}$ has more than one element only in the case where two

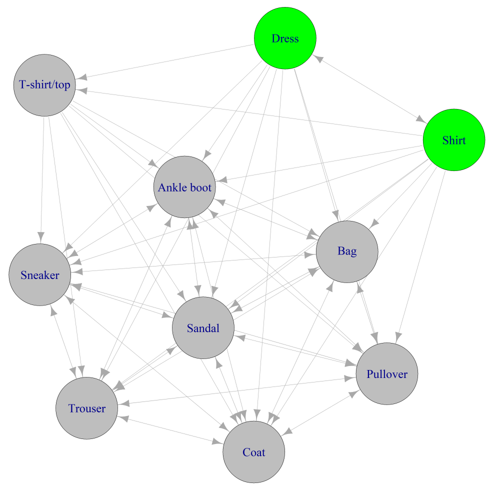

Fig. 1. The directed graph of S_{2pic} associated to m_3.

different classes have the same expected values regarding the *pignistic criterion*. The classifier based on the interval-criterion (2pic) has advantages and disadvantages compared to the *strong dominance* (SD) based one. Indeed, one has to manage cycles in the directed graph due to the intransitivity of I_{2pic}. Note that the combinatorial complexity of the algorithm leading to determine the kernels for S_{2pic} is linear. But the advantage of 2pic compared to SD classifier is that its predictions fall within the subset predicted by SD (see Proposition 4). Therefore, 2pic classifier is more precise than SD and in the same time it is more cautious than the PC classifier.

Proposition 4. *Let us consider a set of finite classes Θ and a mass function m defined over Θ. The following property is verified:*

$$\mathcal{N}_{PC} \subseteq \mathcal{N}_{SD} \text{ and } \mathcal{N}_{PC} \subseteq \mathcal{N}_{2pic}.$$

Proof. Let us consider that $\mathcal{N}_{PC} = \{a\}$, with $a \in \Theta$. If $a \notin \mathcal{N}_{SD}$, then it exists $b \in \Theta$, $b \neq a$ such that $b \succeq_{SD} a$, i.e., $pig_m(b) \geq Bel_m(\{b\}) > Pl_m(\{a\}) \geq pig_m(a)$. Contradiction. If $a \notin \mathcal{N}_{2pic}$, from Definition 12 either $a \notin scck$ or $a \in scck$ but, in both situations, it exists $b \in scck$ such that $(b, a) \in P_{2pic}$. Which means that $Pl_m(\{a\}) < pig_m(b)$. Contradiction.

Proposition 5. *Let us consider a set of finite classes Θ and a mass function m defined over Θ. The following property is verified :*

$\mathcal{N}_{2pic} \subseteq \mathcal{N}_{SD}$ *and it exists some situations where the two subsets are not equal.*

Proof. Let us consider that it exists $a \in \mathcal{N}_{2pic}$ such that $a \notin \mathcal{N}_{SD}$. Then it exists $b \in \mathcal{N}_{SD}$ such that $Bel_m(\{b\}) > Pl_m(\{a\})$ and it follows $pig_m(b) \geq Bel_m(\{b\}) > Pl_m(\{a\})$. Then $b\ P_{2pic}\ a$. This can happen only if a and b belong to the scc containing the kernels of $G = (\Theta, S_{2pic})$ and it exists c different from a and b such that $a\ I_{2pic}\ c$ and $b\ I_{2pic}\ c$. But in the Definition 12 of $\mathcal{N}_{2pic}$, such a class a can not belong to $\mathcal{N}_{2pic}$. Contradiction! Moreover, in the examples of Tables 1 several image predictions give $\mathcal{N}_{2pic} \subset \mathcal{N}_{SD}$.

4 Illustration on Fashion Mnist Data

The illustration of this paper concerns the fashion mnist images. The set-valued measures of performances are calculated on this data for PC, SD and 2pic set-valued classifiers. Note that the mass functions are predicted using the evidential classification based on imprecise relabelling (eclair) method [8] [7] associated with a sequential convolutional neural network (CNN). Fashion-MNIST, a direct drop-in replacement for the original Y. Lecun' MNIST dataset for benchmarking machine learning algorithms [10], is a dataset of Zalando's article images [16] consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 gray-scale image, associated with a label from 10 classes. Note that fashion mnist dataset seems to be more challenging while for mnist dataset classes are clearly separated (see [3]). For the first illustration, we consider only the original training dataset which is split to training (45,000 examples) and validation (15,000 examples) datasets. Here we are interested in the behaviour of the classification based on the PC, SD and 2pic regarding some examples presented in Figure 2. These examples are selected from the validation dataset as they have particular mass functions. Indeed, examples 1 and 2 are selected such that for the class a with maximum belief function value it exists a class b such that $Bel(\{a\}) < Pl(\{b\})$ in a way that a do not dominate b considering the *strong dominance* relation (see Table 1). Consequently, the SD classifier gives a subset of classes as predictions in these situations. While the examples 3 and 4 are selected such that the classes obtaining the maximum value for the pignistic probability and the plausibility function are not the same (see Table 1). In Table 1 it is given the predictions of PC, SD and 2pic. As one can see in

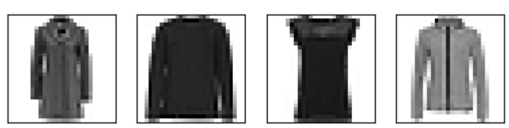

Fig. 2. The fourteen selected fashion images.

Table 1, the results concerning the predictions of 2pic for the first two images are the same as those of PC while SD gives imprecise predictions. For the second two images 2pic gives the same results as the SD classifier while the PC classifier gives wrong predictions. These four selected examples show us some aspects of the behaviour of the 2pic classifier regarding the quality of the data to predict.

Table 1. Bel, Pl, pig and predictions associated to the four examples.

		1 T-shirt/top	2 Trouser	3 Pullover	4 Dress	5 Coat	6 Sandal	7 Shirt	8 Sneaker	9 Bag	10 Ankle boot	truth	pignistic criterion	strong dominance	interval criterion
	Bel_1	0.0	0.0003	0.0096	0.1857	**0.4533**	0.0	0.0319	0.0	0.001	0.0				
$(m_1, \mathbb{F}_1)$	Pl_1	0.0095	0.0099	0.0401	0.4568	**0.7667**	0.0095	0.0797	0.0095	0.0153	0.0095	Coat	Coat	{Dress, Coat}	Coat
	pig_1	0.001	0.0013	0.0192	0.3165	**0.6042**	0.001	0.0508	0.001	0.0042	0.001				
	Bel_2	0.0005	0.0	0.2038	0.0031	0.0703	0.0	**0.419**	0.0	0.0003	0.0				
$(m_2, \mathbb{F}_2)$	Pl_2	0.0087	0.0082	0.5053	0.0113	0.1068	0.0082	**0.7161**	0.0082	0.0085	0.0082	Pullover	Shirt	{Pullover, Shirt}	Shirt
	pig_2	0.0013	0.0008	0.3478	0.0039	0.0818	0.0008	**0.5608**	0.0008	0.0011	0.0008				
	Bel_3	0.214	0.	0.	**0.381**	0.	0.	0.358	0.	0.001	0.				
$(m_3, \mathbb{F}_3)$	Pl_3	0.251	0.002	0.002	0.397	0.002	0.002	**0.403**	0.002	0.003	0.002	Shirt	Dress	{Dress, Shirt}	{Dress, Shirt}
	pig_3	0.231	0.	0.	**0.387**	0.	0.	0.379	0.	0.002	0.				
	Bel_4	0.002	0.	**0.302**	0.002	0.263	0.	0.239	0.	0.	0.				
$(m_4, \mathbb{F}_4)$	Pl_4	0.016	0.013	0.416	0.018	**0.425**	0.013	0.37	0.014	0.013	0.013	Coat	Pullover	{Pullover, Coat, Shirt}	{Pullover, Coat, Shirt}
	pig_4	0.004	0.001	**0.351**	0.005	0.336	0.001	0.297	0.002	0.001	0.001				

It is more cautious than the PC classifier but it bets in more situations than the SD at the risk of being wrong. The second illustration concerns the test data. Table 2 gives the performances of the four classifiers on the test images of the fashion mnist regarding five metrics. The two extremes ones: classical accuracy, imprecision that check if the predicted subset contains the truth and three trade-off well known measures for set-valued classification u_{50}, u_{65} and u_{80} [19]. As one can see, PC classifier has obviously the best accuracy performance as the imprecise predictions are considered wrong for this metric but 2pic has better performances than SD regarding accuracy u_{50}, u_{65} and u_{80} measures. While SD and 2pic have the best performance regarding imprecision metric. The classifier 2pic gives intermediate results between SD and PC for all the metrics. These performances confirm the trade-off behaviour of the classifier 2pic.

Table 2. PC, SD and 2pic performances on the test data of fashion mnist images.

	accuracy	u_{65}	u_{65}	u_{80}	imprecision
PC	**0.916**	**0.916**	**0.916**	0.916	0.916
SD	0.878	0.898	0.905	0.911	**0.936**
2pic	0.883	0.901	0.907	0.913	**0.936**

5 Conclusion

This paper introduce an example of interval criterion-based evidential set-valued Classification to deal with imperfect data in a supervised classification task. The article provides the intuition and the theoretical and experimental justifications of the proposals and their usefulness in terms of controlling the size of the output set, whether the decision maker favors prudence or precision. The proposal of this article is then in line with the idea of promoting techniques of machine learning and AI in general that are reliable and cautious as the decision-maker / user can introduce his preferences to control the output of the algorithms.

Acknowledgement

This paper is supported by the European Union's HORIZON Research and Innovation Programme under grant agreement No 101120657, project ENFIELD (European Lighthouse to Manifest Trustworthy and Green AI).

References

1. Abellan, J., Masegosa, A.R.: Imprecise classification with credal decision trees. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **20**(05), 763–787 (2012)
2. Abellán, J., Moral, S.: Building classification trees using the total uncertainty criterion. International Journal of Intelligent Systems **18**(12), 1215–1225 (2003)
3. Agrawal, A., Ali, A., Boyd, S.: Minimum-distortion embedding. arXiv (2021)
4. Coz, J.J.d., Díez, J., Bahamonde, A.: Learning nondeterministic classifiers. Journal of Machine Learning Research **10**(Oct), 2273–2293 (2009)
5. Denoeux, T.: Decision-making with belief functions: a review. International Journal of Approximate Reasoning **109**, 87–110 (2019)
6. Greco, S., Figueira, J., Ehrgott, M.: Multiple criteria decision analysis, vol. 37. Springer (2016)
7. Imoussaten, A., Jacquin, L.: Cautious classification based on belief functions theory and imprecise relabelling. International Journal of Approximate Reasoning **142**, 130–146 (2022)
8. Jacquin, L., Imoussaten, A., Trousset, F., Montmain, J., Perrin, D.: Evidential classification of incomplete data via imprecise relabelling: Application to plastic sorting. In: Ben Amor, N., Quost, B., Theobald, M. (eds.) Scalable Uncertainty Management. pp. 122–135. Springer International Publishing, Cham (2019)
9. Jacquin, L., Imoussaten, A., Trousset, F., Perrin, D., Montmain, J.: Control of waste fragment sorting process based on mir imaging coupled with cautious classification. Resources, Conservation and Recycling **168**, 105–258 (2021)
10. LeCun, Y.: The mnist database of handwritten digits. http://yann. lecun. com/exdb/mnist/ (1998)
11. Ma, L., Denoeux, T.: Partial classification in the belief function framework. Knowledge-Based Systems p. 106742 (2021)
12. Roy, B.: Classement et choix en présence de points de vue multiples. Revue française d'informatique et de recherche opérationnelle **2**(8), 57–75 (1968)
13. Roy, B., Bouyssou, D.: Aide multicritère à la décision: méthodes et cas. Economica Paris (1993)
14. Troffaes, M.C.: Decision making under uncertainty using imprecise probabilities. International journal of approximate reasoning **45**(1), 17–29 (2007)
15. Vovk, V., Gammerman, A., Shafer, G.: Conformal prediction. Algorithmic learning in a random world pp. 17–51 (2005)
16. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)
17. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on systems, Man, and Cybernetics **18**(1), 183–190 (1988)
18. Zaffalon, M.: Statistical inference of the naive credal classifier. In: ISIPTA. vol. 1, pp. 384–393 (2001)
19. Zaffalon, M., Corani, G., Mauá, D.: Evaluating credal classifiers by utility-discounted predictive accuracy. International Journal of Approximate Reasoning **53**(8), 1282–1301 (2012)