Interpreting Fuzzy Decision Trees with Probability-Possibility Mixtures

Didier Dubois[1], Romain Guillaume[1], Christophe Marsala[2], and Agnès Rico[3]

[1] IRIT-CNRS, Université Jean Jaurès, Toulouse, France
{Didier.Dubois, Romain.Guillaume}@irit.fr
[2] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
Christophe.Marsala@lip6.fr
[3] ERIC, Université Lyon 1, Lyon, France
Agnes.Rico@univ-lyon1.fr

Abstract. The interpretation of the classification result obtained with a fuzzy decision tree is a not-so-easy task as the meaning of the obtained degrees of membership may depend on the type of fuzzy partition involved, ranging from probabilistic to possibilistic readings. Hybrid probabilistic-possibilistic mixtures can provide an interesting way to clarify the underlying components of such a classification result. In this paper, based on the hybrid-mixture model, a new approach is proposed to analyse the result of the classification with a fuzzy decision tree that enhance the explainability of this model.

1 Introduction

The classification of an object with a fuzzy decision tree (FDT) provides a fuzzy subset of the set of labels. The use of this model has been praised highly for its interpretability both in the expression of the knowledge (rules) and also in the way the decision is build from the paths. It is nowadays a commonly used machine learning model in the domain of eXplainable Artificial Intelligence (XAI) [1].

However, even if it is easy to explain the decision provided by an FDT for a given object, its result cannot be interpreted by itself. Indeed, the classification of an object with a FDT produces a fuzzy subset of the set of labels (i.e., decision) that is often interpreted (and thus used) in two opposite ways:

- if the decision must be crisp: the label with the highest membership degree is selected to be the decision associated with the object. Sometimes, the difference between the highest degree and the others is used as a measure of confidence in this decision.
- the decision could be fuzzy: the obtained fuzzy set is kept *as it is* and its analysis provides information about the existence of a possible extra unprecedented decision, not present in the existing set of labels.

Moreover, a strong underlying assumption is made about the result of the classification: in the first case, it is considered as a probability distribution over the set of labels, especially if the notion of fuzzy partition adopted is the one of Ruspini [11]; in the second case, it is considered as a possibility distribution over this set. Even if this assumption is fundamental to the use of a FDT, it is not usually set and it depends on the application domain. Thus, this leads to lessen the interpretability and the justification of the decision given by the tree.

To alleviate this drawback, an interesting task is to offer an approach to better analyse the result of the classification with a FDT and to propose its identification as a probability distribution, a possibility distribution, or, perhaps, a combination of the two. Indeed, such an approach to interpret the classification result of a FDT can bring out an increase of interpretability to this model. To tackle this, we place ourselves in the decision making under uncertainty domain where a decision tree is also a model used to classify alternatives or decisions when the state of affairs is ill-known. Here, the mathematical expectation was proposed by von Neuman and Morgernstern to calculate the utility for probabilistic trees [14]. Dubois and Prade [6] proposed using optimistic and pessimistic possibilistic criteria to calculate utilities of possibilistic lotteries. Both methods satisfy the 3 essential properties required to optimise a decision tree: dynamic consistency, consequentialism and lottery reduction. The last one is important to reduce probability trees to simpler ones and calculate the degree of a class.

Beyond possibilistic and probabilistic mixtures, only a form of hybridisation is possible such that the mixture is possibilistic below a certain threshold and probabilistic above it. This result is presented in [4]. This model, named hybrid probabilistic-possibilistic mixtures, depends on a parameter $\alpha \in [0, 1]$. Recently in [3] a decision model based on probability-possibility mixtures was proposed to evaluate strategies (conditional plans that assign an action to each state where a decision has to be made) in sequential decision making. This model comes down to a convex combination of a possibility distribution and a probability distribution. Moreover in [3] it is presented how to retrieve, in a generally unique way, both distributions from a generalized lottery with weights in $[0, 1]$ whose sum is at least 1.

The aim of this paper is to introduce an approach to analyse the result of the classification with a FDT by means of the use of the hybrid probabilistic-possibilistic mixtures. Thus, a clearer explanation of the decision could be proposed to the user. The paper is organised as follows. Section 2 deals with a brief reminder of fuzzy decision trees. Section 3 is devoted to hybrid probabilistic-possibilistic mixtures and their elicitation. Section 4 presents the hybrid p-π interpretation of FDT. The final Section 5 suggests some future work.

2 Fuzzy decision trees

A fuzzy decision tree is a graphical and hierarchical representation of a fuzzy rule base. Nowadays, it is particularly used in supervised machine learning where it could be built up from a training set summarized as a set of decision rules. Afterwards, it is used for either a characterisation of the training set, or a classification task. Characterisation of the training set is a way to highlight important features, their values and their link to the decision. The classification task aims at assigning a decision (viewed as a class) to any forthcoming example only known by means of its values for each descriptive feature, and whose class is undecided yet.

In the following, after the presentation of fuzzy decision tree as a rule base, we present how the classification is performed with such a tree.

2.1 Fuzzy decision tree as a rule base

A (fuzzy) decision tree is composed of nodes, edges, and leaves: a node is associated with a feature[4]; an edge goes from a node to another node or to a leaf, it is directed from its departure node to its target (node, or leaf) and it is labelled with a particular value, or set of values, of the feature of the departure node; and a leaf is associated with a particular variable, the so-called decision. In a FDT, features and decision are linguistic variables [16], edges are labelled with (fuzzy) labels associated with a membership function, and a leaf is associated with either a crisp or a fuzzy term-set.

A path from the root to a leaf of a FDT is composed of a sequence of nodes linked by edges and it is equivalent to an IF...THEN rule. The premises for such a rule r are composed of values of features, and the conclusion is the value that labels the leaf of the path. For instance, a rule with p premises:

`if` i_1 `is` v_{i_1} `and ...and` i_p `is` v_{i_p} `then the decision is` y_r,

with $i_1, \ldots, i_p$ several linguistic variables, v_j a particular value of the linguistic variable j associated with the membership function μ_{v_j}, and y_r a particular value of the decision associated with the membership function μ_{y_r}. Membership degrees are valued independently at each node, they only depend on the values of object $\mathbf{x}$ for the features present in the nodes.

Sometimes, the decision is associated with several values, in this paper, for the sake of clarity, we focus on the case where a decision is only associated with a unique value.

In some contexts, a FDT can be described by a human expert according to her/his knowledge, but more commonly, the FDT is automatically built up from a set of objects. Indeed, in supervised machine learning, FDT is built up from a training set $\mathcal{X}$ that is a set of n objects $\mathbf{x}$ described by means of d features from $\mathcal{F} = \{1, 2, \ldots, d\}$ and associated with a label (or class) from a set Y [9,13,15]. In our context, features are linguistic variables, and terminal labels are linguistic values. Each label corresponds to a decision. The fuzzy partitions of the linguistic variables can be either provided as domain knowledge to the FDT construction algorithm, as for instance in [15], or they can be automatically built from training set during the learning process, as for instance in [9].

In the first case, it is usual to define a Ruspini fuzzy partition [11] associated with the linguistic variables since such a kind of partitions ensures a good explainability [1] of the FDT for the user, particularly in monotone fuzzy decision tree as it is shown in [10]. In the second case, the fuzzy partition could be of any

[4] Features are sometimes called variables, or attributes according to the domain.

kind, satisfying only the constraints that the intersection of the cores of any pair of fuzzy sets is empty and the union of their supports covers the whole universe of values of the variable (e.g., as in [7]).

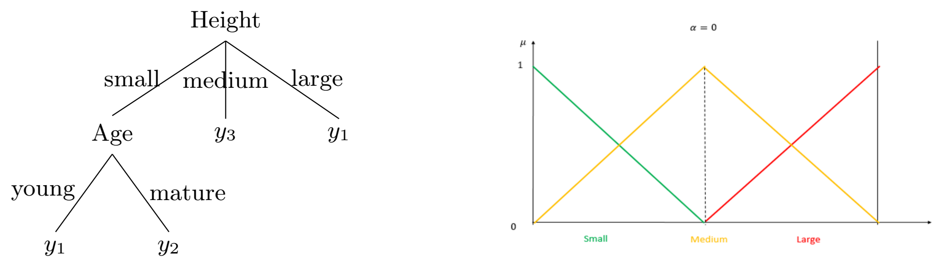

Fig. 1. Fuzzy decision tree (left) and Ruspini fuzzy partition for feature Height (right)

Example 1. Let $\mathcal{F}$ be the set $\{Height, Age\}$ of features, with {*small, medium, large*} the values of the feature *Height* and {*young, mature*} the values of *Age*. An instance of fuzzy decision tree on $\mathcal{F}$ with a Ruspini partition on the variable *Height* are shown in Fig. 1. This FDT is equivalent to a fuzzy rule base with 4 rules, each one associated with a path in the tree. For instance, the path on the left of the tree, leading from the root note *Height* to the leaf that contain the label y_1 gives the following rule:
`if` ***`Height`*** `is` ***`small`*** `and` ***`Age`*** `is` ***`young`*** `then the decision is` y_1. ○

2.2 Classification with a fuzzy decision tree

Let $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ be an object to be classified by means of a fuzzy decision tree $\mathcal{T}$. This object is described by means of d values each associated with a particular feature. The matching degree of $\mathbf{x}$ with the path r of the tree with p_r premises is valued with a t-norm $\top$:

$$\mu_r(\mathbf{x}) = \top_{j \in \{i_1, \ldots, i_{p_r}\}} \mu_{v_j}(x_j).$$

The path r is leading to a leaf associated with a label $y \in \mathcal{Y}$, as a consequence, $\mathbf{x}$ is associated with the label y with the degree $\mu_r(\mathbf{x})$ according to r.

All the K leaves could be reached during the classification of $\mathbf{x}$, each label present in a leaf is thus associated with a degree related to its path. To obtain the global membership degree of $\mathbf{x}$ to each label $y \in \mathcal{Y}$, an aggregation of all these degrees by means of a t-conorm is used:

$$\forall y \in \mathcal{Y},\ \mu_y(\mathbf{x}) = \bot_{\{r \in \{1, \ldots K\} \mid y_r = y\}} \mu_r(\mathbf{x}).$$

Thus, we have: $\forall y \in \mathcal{Y},\ \mu_y(\mathbf{x}) = \bot_{\{r \in \{1, \ldots K\} \mid y_r = y\}} \top_{j \in \{i_1, \ldots, i_{p_r}\}} \mu_{v_j}(x_j)$.

Usually, following Zadeh, the minimum t-norm and the maximum t-conorm are chosen to compute these membership degrees:

$$\forall y \in \mathcal{Y},\ \mu_y(\mathbf{x}) = \max_{\{r\in\{1,\dots K\}\ |\ y_r=y\}} \min_{j\in\{i_1,\dots,i_{p_r}\}} \mu_{v_j}(x_j). \qquad (1)$$

Other pair of dual t-norm-t-conorm can be used, for instance, Łukasiewicz or probabilistic ones. If Ruspini partitions are used, choosing the product for $\top$ and the sum for $\bot$, then for each input $\mathbf{x}$, the FDT turns into a probability tree, the distribution of weights $(\mu_{y_1}(\mathbf{x}),\dots,\mu_{y_R}(\mathbf{x}))$ is such that $\sum_{r=1}^{R}\mu_{y_r}(\mathbf{x}) = 1$ and can be interpreted as a probability distribution.

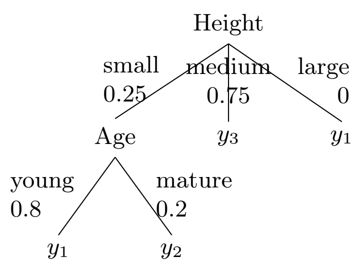

Fig. 2. A fuzzy decision tree on $\mathcal{F}$

Example 2. We consider the FDT presented in Example 1 and shown in Fig. 2. The classification are based on the minimum t-norm and maximum t-conorm. The object $\mathbf{x}$, with value x_H for feature Height and x_A for Age, to classify has been compared to all nodes. We found, for all edges, the membership degrees of $\mathbf{x}$ recalled in Fig. 2: $\mathbf{x}$ has

• a Height x_H that is small with a degree $\mu_{sm}(x_H) = 0.25$, medium with a degree $\mu_{me}(x_H) = 0.75$,

• an Age x_A that is young with a degree $\mu_{yo}(x_A) = 0.8$ and mature with a degree $\mu_{ma}(x_A) = 0.2$.

According to the tree, $\mathbf{x}$ is associated to the leaf of the first path with the degree $\mu_{r1}(\mathbf{x}) = \min(\mu_{sm}(x_A), \mu_{yo}(x_H)) = 0.25$, thus $\mathbf{x}$ is associated to label y_1 (the label in the leaf) with the membership degree $\mu_{r1}(\mathbf{x})$. In the same way, we value the membership degree of $\mathbf{x}$ to the other path: $\mu_{r2}(\mathbf{x}) = 0.2$, $\mu_{r3}(\mathbf{x}) = 0.75$, $\mu_{r4}(\mathbf{x}) = 0$. Paths $r1$ and $r4$ lead to label y_1, thus the membership degree of $\mathbf{x}$ to y_1 is valued as $\mu_{y_1}(\mathbf{x}) = \max(\mu_{r1}(\mathbf{x}), \mu_{r4}(\mathbf{x})) = \max(0.25, 0) = 0.25$. Similarly, $\mathbf{x}$ is associated to y_2 with the degree $\mu_{y_2}(\mathbf{x}) = 0.2$, and to y_3 with the degree $\mu_{y_3}(\mathbf{x}) = 0.75$.

Summing up, the label associated to $\mathbf{x}$ with a classification with the FDT is the fuzzy set $0.25|y_1 + 0.2|y_2 + 0.75|y_3$. If a decision has to be taken between one of the labels, then label y_3 could be chosen as it is associated with the highest membership degree. It could eventually be associated with a confidence degree valued by means of the obtained membership degree of y_3 and the other ones. For instance, the confidence degree could be valued as $0.75 - \max(0.25, 0.2) = 0.5$.

If the decision could be handled as it is, an analysis of the fuzzy set could be performed. Here, it highlights the fact that $\mathbf{x}$ is *in an area* in-between the 3 classes, strongly in y_3, and also close to y_1 and y_2. Possibly, a new kind of decision, combining y_1, y_2, and y_3 could be drawn. To illustrate this, let the labels be diseases, with y_1 the flu, y_2 the bronchitis, and y_3 the pneumonia. In the first case, the decision leads to predict that patient $\mathbf{x}$ has a pneumonia with a good confidence degree. In the second case, the decision may highlight a possible new kind of disease never identified in the past. ◦

In the following section, the model of hybrid probabilistic-possibilistic mixtures is applied to the classification result of the FDT to better interpret it.

3 Hybrid probabilistic-possibilistic mixtures

In this section we present the hybrid probabilistic-possibilistic mixtures ("p-π mixtures" for short) introduced in [4] and their elicitation procedure introduced in [3]. The aim of p-π mixtures is to represent a set of uncertainty values as a convex combination of a probability distribution and a possibility distribution, a weighted average that weights each of the two distributions so as to lay bare its importance within the original distribution. As a consequence, it becomes possible to understand clearly what is the underlying nature of the uncertainty disptribution, providing a better explainability. Hereafter, we first introduce the basis of p-π mixtures, and secondly we present their elicitation.

3.1 Definition and properties of the p-π mixtures

Let $X = \{1, \cdots, m\}$ be a finite set and S be a t-conorm on $[0, 1]$. S is a binary operation $S : [0, 1]^2 \to [0, 1]$ such that for all $x, y, z \in [0, 1]$: S is commutative, associative, increasing according to both arguments, and satisfies the boundary condition $S(x, 0) = x$. In the following, we consider non additive measures $\rho : 2^X \to [0, 1]$ that are S-decomposable [5], i.e., such that $\rho(\emptyset) = 0$, $\rho(X) = 1$ and for all disjoint $A, B \subseteq X$, $\rho(A \cup B) = S(\rho(A), \rho(B))$.

To be used as in probability theory, this model needs to offer a counterpart of the notion of probabilistic independence. So the following property is defined: two events $A, B \subseteq X$ are called $\star$-separable with respect to a decomposable non-additive measure ρ if there exists a t-norm $\star$ such that $\rho(A \cap B) = \rho(A) \star \rho(B)$. In probability theory, if each of two disjoint events $A, B \subseteq X$ are independent of another event C, then $A \cup B$ is independent from C as well. This property is essential for the reduction of probability trees into simple probability distributions. Enforcing this property on S-decomposable non-additive measures leads to require the following distributivity property [4]: if $A \cap B = \emptyset$,

$$\rho(A \cup B) \star \rho(C) = S(\rho(A), \rho(B)) \star \rho(C) = S(\rho(A) \star \rho(C), \rho(B) \star \rho(C)) \quad (2)$$

since $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. The only (t-conorm, t-norm) pairs that satisfy this property are known to be such that [8]: $\exists \alpha \in [0, 1], \forall x, y \in [0, 1]$,

$$S^\alpha(x,y) = \begin{cases} \min(1, x+y-\alpha) \text{ if } x > \alpha \text{ and } y > \alpha, \\ \max(x,y) \text{ otherwise.} \end{cases} \quad (3)$$

$$x \star_\alpha y = \begin{cases} \alpha + \frac{(x-\alpha)(y-\alpha)}{1-\alpha} & \text{if } x > \alpha \text{ and } y > \alpha, \\ x \top y & \text{otherwise,} \end{cases} \quad (4)$$

where $\top$ is a triangular norm. Note that we have $S^\alpha(x,y) = \max(x, y, \min(1, x + y - \alpha))$. In the following we use the minimum operation for $\top$. Note that the t-conorm S^α and t-norm $\star_\alpha$ are not dual each other.

Decomposable non-additive measures allowing for the reduction of probability trees, i.e., satisfying the property (2) are denoted by ρ^α such that, as explained in [4],

$$\rho^\alpha(A) = S^\alpha_{x_i \in A} \ \rho^\alpha_i \quad (5)$$

where $\rho^\alpha_i = \rho^\alpha(\{x_i\})$, since S^α is associative.

It turns out that set-functions ρ^α can be rewritten as Shafer's plausibility functions [12] of the form $\rho^\alpha(A) = \alpha\Pi(A) + (1-\alpha)P(A)$, where Π is a possibility measure, P is a probability measure, and α is a fixed parameter. In particular, $\rho^\alpha_i = \alpha\pi_i + (1-\alpha)p_i$. Moreover, the associated distributions p and π must satisfy the constraint $p_i = 0$ if $\pi_i < 1$ (see [2] for more details). Note that ρ^α is a possibility distribution if $\alpha = 1$ and a probability distribution if $\alpha = 0$. Intuitively, the interpretation is that we have probabilities only on the completely possible states. To be more precise we have the following equivalences.

- $\rho^\alpha_i > \alpha$ is equivalent to $p_i > 0$ and these conditions imply $\pi_i = 1$.
- $\rho^\alpha_i = \alpha$ is equivalent to $\pi_i = 1$ and $p_i = 0$.
- $\rho^\alpha_i < \alpha$ is equivalent to $\pi_i < 1$ and $p_i = 0$.

We conclude this part with the normalisation condition that the coefficients ρ^α_i must satisfy. We denote C^+_α the set $\{i : \rho^\alpha_i > \alpha\} = \{i : p_i > 0\}$. If $C^+_\alpha \neq \emptyset$, the normalisation condition $\rho^\alpha(X) = 1$ takes the form: $\sum_{i \in C^+_\alpha} \rho^\alpha_i - \alpha(|C^+_\alpha| - 1) = 1$. This condition can also be written: $\sum_{i=1}^m \max(0, \rho^\alpha_i - \alpha) = 1 - \alpha$. Note that:

- If $\alpha = 0$ (ρ^0 is a probability measure), the normalisation condition becomes $\sum_{i=1}^m \rho^0_i = 1$.
- If $\alpha = 1$, (ρ^1 is a possibility measure), the normalisation condition becomes $\max_{i=1}^m \rho^1_i = 1$.

3.2 Elicitation of p-π mixtures

This section is a brief reminder of the result presented in [3] where the hybrid model p-π is used to interpret a distribution of weights $\rho = (\rho_1, \dots \rho_m) \in [0,1]^m$ with $m \geq \sum_{i \in [m]} \rho_i \geq 1$. We assume that ρ satisfies neither the normalisation condition of probability distributions nor the one for possibility distributions. Usually, two normalisation options could be investigated:

1. dividing the weights by their sum to build a probability distribution;
2. dividing the weights by their maximum to make a possibility distribution.

Here, we are going to interpret a distribution of weights as an hybrid p-π distribution without altering this distribution. As shown in [3], for any distribution of weights $\rho = (\rho_1, \cdots, \rho_m) \in [0, 1[^m$, there exists a unique parameter value $\alpha \in [0, 1]$, a unique possibility distribution π and a unique probability distribution p such that $\rho_i = \alpha\pi_i + (1 - \alpha)p_i$ for all $i \in X$. This distribution is denoted by ρ_i^α and it is calculated as follows. We consider $n \leq m$ such that $\rho_{(n)} < \cdots < \rho_{(1)}$ with $R_{(i)} = \{j|\rho_j = \rho_{(i)}\}$. Note that if $\alpha \in [\rho_{(i+1)}, \rho_{(i)})$ then $\{i : \rho_i > \alpha\} = \cup_{j=1}^{i} R_{(j)}$. We have the following result.

Theorem 1 ([3]). *We consider m weights $\rho_j \in [0, 1], j = 1, \ldots, m$ such that $\sum_{j=1}^{m} \rho_j \geq 1$ and $\max_{j=1}^{m} \rho_j < 1$. There exists a unique value α, an integer i_0 such that $\rho_{(i_0+1)} \leq \alpha < \rho_{(i_0)}$, a unique possibility distribution π and a unique probability distribution p such that $\rho_i = \alpha\pi_i + (1 - \alpha)p_i$ $\forall i$ with*

- *α solution of $\sum_{i=0}^{m} \max(0, \rho_i^\alpha - \alpha) = 1 - \alpha$.*
- *$\forall i \geq i_0 + 1$ $\forall j \in R_{(i)}, \pi_j = \frac{\rho_j}{\alpha}$ and $p_j = 0$*
- *$\forall i \leq i_0$ $\forall j \in R_{(i)}, p_j = \frac{\rho_j - \alpha}{1 - \alpha}$ and $\pi_j = 1$.*

Note that if $\max_{j=1}^{m} \rho_j = \rho_{(1)} = 1$, and $R_{(1)} = \{i^*\}$ then any $\alpha \in [\rho_{(2)}, 1]$ may be chosen and we have $\rho_i = \alpha\pi_i + (1 - \alpha)p_i$ with $p_{i*} = 1$ and $\pi_j = \rho_j/\alpha$, for $j \neq i^*$.

4 Hybrid p-π mixtures for fuzzy decision trees

This section presents how to use the hybrid p-π mixtures in order to obtain a p-π decision tree. The aim is to decompose a FDT as a combination of a probabilistic tree and a possibilistic tree when it is used to classify an object. To this end, we are going to use the above elicitation of a p-π mixture.

4.1 The model

We consider an example $\mathbf{x} = (x_1, \cdots, x_n)$ to be classified using a FDT. The object $\mathbf{x}$ is compared to all nodes and we find, for all edges, the membership degrees of $\mathbf{x}$ according to the possible fuzzy classes.

For instance let us consider a feature $F_k \in \mathcal{F}$ with its linguistic referential $L = \{l_{k1}, l_{k2}, ...\}$. Object $\mathbf{x}$ has an evaluation x_k according to F_k. We get some membership degree on each edge denoted by $\rho_{l_{ki}}(x_k) \in [0, 1]$, for all $l_{ki} \in L$. We consider all the degrees involved in the FDT, and we denote η the number of its edges and $[\eta]$ the set $\{1, \cdots, \eta\}$. Edges are numbered such that the degrees ρ_k, $k \in [\eta]$ are in decreasing order. We apply Theorem 1 to elicit the parameter α, the probability p and the possibility measure π over all edges.

The matching degree of $\mathbf{x}$ with the path r of the tree with p_r premises is valued with $\star_\alpha$. In order to have friendly notation, we denote the vertices of the path r by $I = \{i_1, \cdots, i_{k_r}\}$, and $\rho_{i_1}, \cdots, \rho_{i_{k_r}}$ are the degrees present in the path. Hence the matching degree of $\mathbf{x}$ with the path r is:

$$\rho_r(\mathbf{x}) = \star_{\alpha\, i \in I} \rho_{l_i}(x_i).$$

The path r is leading to a leaf y. The global membership degree of $\mathbf{x}$ to each label y is an aggregation of all these degrees $\rho_r(\mathbf{x})$ for all path r arriving to the leaf y. We denote this set of path by R_y.The aggregation is made with S^α. One obtains the global membership degree of $\mathbf{x}$ to y as follows:

$$\rho_y(\mathbf{x}) = S^\alpha_{r \in R_y} \rho_r(\mathbf{x}).$$

From equations (3) and (4), it is worth noticing that if for each path at least one membership degree is less than parameter α the presented method is the classical one involving the minimum t-norm and the maximum t-conorm.

4.2 Examples

This section presents examples that illustrate the various cases we can encounter. The first one is equivalent to the classical aggregation with the minimum and the maximum as t-norm and t-conorm respectively. The second one has degrees greater than the parameter α. The last one considers a weight 0.5 at each node of the FDT.

Example 3. We use again the FDT presented in Fig. 1, where $\mathcal{F}$ is the set of features $\{Height, Age\}$, with {small, medium, large} the linguistic referential of the feature *Height* and {young, mature} the one of *Age*. Let $\mathbf{x} = (x_H, x_A)$ be the example to classify, with its membership degrees to each feature terms valued according to a generalised (i.e. non-Ruspini) partition (the sum of all degrees is not equal to 1): the Height x_H of $\mathbf{x}$ is small with a degree $\rho_{sm}(x_H) = 0.25$, medium with a degree $\rho_{me}(x_H) = 0.4$, and large with a degree $\rho_{la}(x_H) = 0.75$. The Age x_A of $\mathbf{x}$ is young with a degree $\rho_{young}(x_A) = 0.8$, and mature with a degree $\rho_{ma}(x_A) = 0.2$. All these degrees are involved in the FDT. Edges are numbered such that the degrees are in decreasing order. So one obtains $\rho_1 = 0.8, \rho_2 = 0.75, \rho_3 = 0.4, \rho_4 = 0.25$ and $\rho_5 = 0.2$.
Let us calculate α such that $\rho = (\rho_1, \rho_2, \rho_3, \rho_4, \rho_5)$ is an α p-π mixture.

- Let us suppose that $i = 1$, $\alpha \in [0.75, 0.8)$. Then $C^+_\alpha = \{1\}$.
  The normalisation condition is $\sum_{i \in C^+_\alpha} \rho_i^\alpha - \alpha(|C^+_\alpha| - 1) = 0.8 \neq 1$ so $i \neq 1$.
- Let us suppose that $i = 2$, $\alpha \in [0.4, 0.75)$. Then $C^+_\alpha = \{1, 2\}$.
  The normalisation condition is $0.8 + 0.75 - \alpha = 1$ i.e. $\alpha = 0.55 \in [0.4, 0.75)$

So $\alpha = 0.55$ is the solution and $i_0 = 2$.

- $\forall j \geq 3$ $\pi_j = \frac{\rho_j}{\alpha}$ and $p_j = 0$
- $\forall j \leq 2$ $p_j = \frac{\rho_j - \alpha}{1 - \alpha}$ and $\pi_j = 1$.

i	1	2	3	4	5
π_i	1	1	$\frac{0.4}{0.55}$	$\frac{0.25}{0.55}$	$\frac{0.2}{0.55}$
p_i	$\frac{0.25}{0.45}$	$\frac{0.2}{0.45}$	0	0	0

i	1	2	3	4	5
π_i	1	1	1	1	$\frac{0.5}{0.76}$
p_i	0.5	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0

Table 1. Elicitation of p and π in Example 3 (left) and Example 4 (right)

We have $\rho_i = 0.55\pi_i + 0.45p_i$ and the degrees of the different classes y are calculated using $S^{0.55}$ and $\star_{0.55}$ defined as follows

$$S^{0.55}(x,y) = \begin{cases} \min(1, x+y-0.55) \text{ if } x > 0.55, y > 0.55 \\ \max(x,y) \text{ otherwise,} \end{cases}$$

$$x \star_{0.55} y = \begin{cases} 0.55 + \frac{(x-0.55)(y-0.55)}{0.45} & \text{if } x > 0.55 \text{ and } y > 0.55 \\ \min(x,y) & \text{otherwise} \end{cases}$$

We obtain the following degrees for classes y_1, y_2 and y_3: the degree for y_3 is 0.4, the degree for y_2 is $0.25 \star_{0.55} 0.2 = 0.2$, and the degree for y_1 is $S^{0.55}(0.75, 0.25 \star_{0.55} 0.8) = S^{0.55}(0.75, 0.25) = 0.75$. The complete elicitation of p and π is shown in Table 1. Note that, in each calculation, one of the degrees is less than 0.55, so the results are similar with Zadeh's $\max - \min$ connectives. ○

Example 4. As in the previous case, we use the FDT presented in Fig. 1 with different membership degrees. The Height x_H of $\mathbf{x}$ is small with a degree $\rho_{sm}(x_H) = 0.8$, medium with a degree $\rho_{me}(x_H) = 0.8$, and large with a degree $\rho_{la}(x_H) = 0.9$. The Age x_A of $\mathbf{x}$ is young with a degree $\rho_{young}(x_A) = 0.8$, and mature with a degree $\rho_{ma}(x_A) = 0.2$. We consider all degrees involved in the FDT. Edges are numbered such that the degrees ρ_k, $k \in [\eta]$, are in decreasing order. So one obtains $\rho_1 = 0.9, \rho_2 = 0.8, \rho_3 = 0.8, \rho_4 = 0.8$ and $\rho_5 = 0.2$.
Let us calculate α such that ρ is a p-π mixture.
Let us suppose that $i = 1$, $\alpha \in [0.8, 0.9)$. The normalisation condition is $\sum_{i \in C_\alpha^+} \rho_i^\alpha - \alpha(|C_\alpha^+| - 1) = 0.9 + 0 \neq 1$, as a consequence $i \neq 1$.
Let us suppose that $i = 2$, $\alpha \in [0.2, 0.8)$. The normalisation condition is $2.4 + 0.9 - 3\alpha = 1$ i.e. $\alpha = 0.76 \in [0.2, 0.8)$. So $\alpha = 0.76$ is the solution and $i_0 = 2$ and we have $\forall j \geq 3$, $\pi_j = \frac{\rho_j}{\alpha}$ and $p_j = 0$; and $\forall j, \leq 2$ $p_j = \frac{\rho_j - \alpha}{1-\alpha}$ and $\pi_j = 1$.
We have $\rho_i = 0.76\pi_i + 0.24p_i$ and the degrees of the classes y are calculated using $S^{0.76}$ and $\star_{0.76}$ defined as follows

$$S^{0.76}(x,y) = \begin{cases} \min(1, x+y-0.76) \text{ if } x > 0.76, y > 0.76 \\ \max(x,y) \text{ otherwise,} \end{cases}$$

$$x \star_{0.76} y = \begin{cases} 0.76 + \frac{(x-0.76)(y-0.76)}{0.24} & \text{if } x > 0.76 \text{ and } y > 0.76 \\ \min(x,y) & \text{otherwise} \end{cases}$$

We obtain the following degrees for the classes y_1, y_2 and y_3: the degree for y_3 is 0.8, the degree for y_2 is 0.2 and the degree for y_1 is 0.91. The complete elicitation of p and π is shown in Table 1. ○

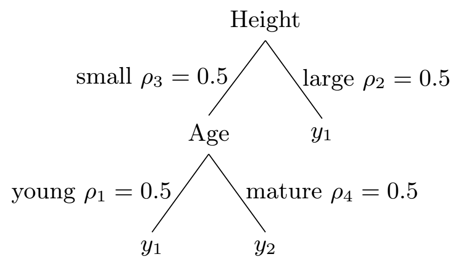

Fig. 3. A completely uniform FDT

Example 5. We consider the FDT presented in Fig. 3. If $\alpha < 0.5$, then the normalisation condition is $4(0.5 - \alpha) = 1 - \alpha$ i.e., $\alpha = \frac{1}{3}$. We have $\frac{1}{3} < 0.5$ so $\alpha = \frac{1}{3}$ is accepted. For all i we have $\rho > \alpha$ so for all i, $\pi_i = 1$ and $p_i = \frac{0.5-\frac{1}{3}}{1-\frac{1}{3}} = \frac{1}{4}$. So we have $\pi_1 = \pi_2 = \pi_3 = \pi_4 = 1$ and $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$ with $\alpha = \frac{1}{3}$.

We have $\rho_i = \frac{1}{3}\pi_i + \frac{2}{3}p_i$ and the degrees of the different class y are calculated using $S^{\frac{1}{3}}$ and $\star_{\frac{1}{3}}$ defined as follows

$$S^{\frac{1}{3}}(x, y) = \begin{cases} \min(1, x + y - \frac{1}{3}) \text{ if } x > \frac{1}{3}, y > \frac{1}{3} \\ \max(x, y) \text{ otherwise,} \end{cases}$$

$$x \star_{\frac{1}{3}} y = \begin{cases} \frac{1}{3} + \frac{(x-\frac{1}{3})(y-\frac{1}{3})}{\frac{2}{3}} & \text{if } x > \frac{1}{3} \text{ and } y > \frac{1}{3} \\ \min(x, y) & \text{otherwise} \end{cases}$$

We obtain the degree $\frac{13}{24}(> 0.5)$ for leaf y_1, and the degree $\frac{3}{8}(< 0.5)$ for leaf y_2. Even with uniform FDT, this does not lead to a uniform distribution of degrees across all classes as y_1 is present in a greater number of leaves than y_2. ◦

From the above examples, it is clear that the larger the weights ρ_i in the fuzzy decision tree, the closer we get to a possibilistic interpretation of the degrees of membership. Another interpretation concerns the degrees calculated for the leaves. These degrees are between those obtained with possibilistic trees and those obtained with probabilistic trees.

5 Conclusion and future work

In this paper, a new approach is introduced to analyse the result of the classification with a FDT in order to increase its explainability. This approach is based on the use of the hybrid probabilistic-possibilistic mixtures and provides a new way to offer an explanation of the decision to the user and help her/him to choose the best use of the classification result. Indeed, this approach offers an interpretation of the result of the classification with a FDT as a probability distribution, a possibility distribution, or a combination of the two. We focus on an interpretation with the hybrid p-π mixture that is done globally, with the consideration of the membership degrees of all edges as a whole.

In future work, an interpretation of fuzzy labels as p-π mixtures will be investigated. The question is then how to compute these fuzzy labels according to the choice of the (t-conorm, t-norm) pair. Another perspective is a layered approach where the interpretation is made separately at each node. The question is then how to aggregate various α's that might appear. A last perspective is to define a family of partitions that itself satisfy the normalisation condition for a given α as the Ruspini partition satisfy the probability normalisation.

References

1. J. M. Alonso Moral, C. Castiello, L. Magdalena, and C. Mencar. *Explainable fuzzy systems*. Studies in Computational Intelligence. Springer Cham, 2021.
2. D. Dubois, H. Fargier, R. Guillaume, and A. Rico. Sequential decision-making under uncertainty using hybrid probability-possibility functions. In V. Torra and Y. Narukawa, editors, *Int. Conf. on Modeling Decision for Artificial Intelligence – MDAI*, volume 12898 of *LNCS*, pages 54–66, Umeå, Sweden, 2021. Springer.
3. D. Dubois, R. Guillaume, and A. Rico. Eliciting hybrid probability-possibility functions and their decision evaluation models. In E. Miranda, I. Montes, E. Quaeghebeur, and B. Vantaggi, editors, *ISIPTA 2023*, volume 215 of *Proc. of Machine Learning Research*, pages 200–209, Oviedo, Spain, 2023.
4. D. Dubois, E. Pap, and H. Prade. Hybrid probabilistic-possibilistic mixtures and utility functions. In János Fodor, Bernard De Baets, and Patrice Perny, editors, *Preferences and Decisions under Incomplete Knowledge*, pages 51–73. Physica-Verlag HD, Heidelberg, 2000.
5. D. Dubois and H. Prade. A class of fuzzy measures based on triangular norms. *Int. J. of General Systems*, pages 43–61, 1982.
6. D. Dubois and H. Prade. Possibility theory as a basis for qualitative decision theory. *IJCAI 1995*, 95:1924–1930, 1995.
7. D. Dubois and D. Sánchez. Fuzzy clustering based on coverings. In C. Borgelt, M.Á. Gil, J.M.C. Sousa, and M. Verleysen, editors, *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*, pages 319–330. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
8. E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Acad., 2000.
9. C. Marsala and B. Bouchon-Meunier. An adaptable system to construct fuzzy decision trees. In *NAFIPS 1999*, pages 223–227, New York, USA, June 1999.
10. C. Marsala and D. Petturiti. Splitting rules for monotone fuzzy decision trees. In S. Massanet, S. Montes, D. Ruiz-Aguilera, and M. González-Hidalgo, editors, *Fuzzy Logic and Technology, and Aggregation Operators*, pages 161–173. Springer Nature Switzerland, 2023.
11. E. H. Ruspini. A new approach to clustering. *Inf. and Control*, 15(1):22–32, 1969.
12. G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.
13. M. Umano, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu, and J. Kinoshita. Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems. In *Fuzz-IEEE*, volume 3, pages 2113–2118, Florida, june 1994.
14. J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
15. Y. Yuan and M.J. Shaw. Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, 69:125–139, 1995.
16. L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning – I. *Information Sciences*, 8(3):199–249, 1975.