Representation of Explanations of Possibilistic Inference Decisions

Ismaïl Baaj, Jean-Philippe Poli, Wassila Ouerdane, Nicolas Maudet

Representation of Explanations of Possibilistic Inference Decisions

Ismaïl Baaj
Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
LIP6, Sorbonne Université, Paris, France
ismail.baaj@cea.fr,ismail.baaj@lip6.fr

Jean-Philippe Poli
Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France
jean-philippe.poli@cea.fr

Wassila Ouerdane
MICS, CentraleSupélec, Université Paris-Saclay, Gif sur Yvette, France
wassila.ouerdane@centralesupelec.fr

Nicolas Maudet
LIP6, Sorbonne Université, Paris, France
{nicolas.maudet}@lip6.fr

Abstract

In this paper, we study how to explain to end-users the inference results of possibilistic rule-based systems. We formulate a necessary and sufficient condition for justifying by a relevant subset of rule premises the possibility degree of each output attribute value. We apply functions to reduce the selected premises, in order to form two kinds of explanations: the *justification* and the *unexpectedness* of the possibility degree of an output attribute value. The justification is composed of possibilistic expressions that are sufficient to justify the possibility degree of the output attribute value. The unexpectedness is a set of possible or certain possibilistic expressions, which are not involved in the determination of the considered inference result although there may appear to be a potential incompatibility between them and the considered inference result.

We then define a representation of explanations of possibilistic inference decisions that relies on conceptual graphs and may be the input of natural language generation systems. Our extracted justification and unexpectedness are represented by nested conceptual graphs. All our constructions are illustrated with an example of a possibilistic rule-based system that controls the blood sugar level of a patient with type 1 diabetes.

1 Introduction

Possibility Theory is a well-known framework for the handling of incomplete or imprecise information [7, 8] that models the uncertainty by two dual measures called possibility and necessity. These measures allow to distinguish between what is possible without being certain at all and what is certain to some extent. The possibilistic handling of rule-based systems [10, 12] has led to the emergence of possibilistic rule-based systems used for medical diagnostics [4] e.g., DIABETO [16]. For safety-critical applications such as medicine, the generation of explanations of the decisions made by AI systems is now a legitimate demand, in view of the recent adoption of laws that reinforce users' rights e.g., GDPR [13]. Recently, an emphasis was put on possibilistic rule-based systems [9], where the authors highlighted the approach of [11] to develop the explanatory capabilities of these systems. In [11], Farreny and Prade propose to perform a sensitivity analysis by using a min-max equations system. By an example, the authors suggest that it is possible to justify an inference result by some rule premises. They also give a natural language explanation of an inference result of their example. In fact, their approach aims at generating explanations of possibilistic inference decisions, which have to be expressed in natural language for end-users. To generate them with Natural Language Generation techniques [14], authors of [2] propose to define a representation of the explanations of inference decisions. In this paper, we elaborate the explanatory capabilities of possibilistic rule-based systems. Our purpose is twofold. First, we study how to select rule premises justifying their inference results. Then, we define a graphical representation of explanations constructed from the selected premises. We first remind the inference mechanism of a possibilistic rule-based system, introduce useful notations and give an example of such a system, which will be used to illustrate all the constructions of the paper (section 2). The inference result of a possibilistic rule-based system is an output possibility distribution, which assigns to each output attribute value a possibility degree. In Section 3, we give a necessary and sufficient condition to justify by rule premises the possibility degree of any output attribute value. Under this condition, we extract a corresponding subset of premises.

In Section 4, we define four premise reduction functions and apply them to the subset of premises of section 3. This leads us to form two kinds of explanations: the justification and the unexpectedness of the considered output attribute value. The justification is formed by reducing the selected premises to the structure responsible for their possibility or necessity degree. It uses two premise reduction functions. The unexpectedness is a set of possible or certain possibilistic expressions related to the considered inference result in the following sense: although there may appear to be a potential incompatibility between each of the possibilistic expressions and the considered inference result, they are not involved in the determination of the inference result. They are extracted by applying the two other premise reduction functions.

We then propose to represent explanations by conceptual graphs, which provide a natural way to represent knowledge by concepts and n-ary relations (section 5).

Using our justification and unexpectedness, we represent explanations by nested conceptual graphs. Finally, in section 6, we conclude with some perspectives.

2 Background

In this section, following [9], we remind the inference mechanism of possibilistic rule-based systems. Some notations which will be useful in the rest of the paper are introduced. We also give an example of a possibilistic rule-based system.
We consider a set of n parallel if-then possibilistic rules $R^1, R^2, \cdots, R^n$, where each R^i is of the form: "if p_i then q_i" and has its uncertainty propagation matrix $\begin{bmatrix} \pi(q_i|p_i) & \pi(q_i|\neg p_i) \\ \pi(\neg q_i|p_i) & \pi(\neg q_i|\neg p_i) \end{bmatrix} = \begin{bmatrix} 1 & s_i \\ r_i & 1 \end{bmatrix}$. The premise p_i of the rule R^i is of the form $p_i = p_1^i \wedge p_2^i \wedge \cdots \wedge p_k^i$, where each p_j^i is a proposition: "$a_j^i(x) \in P_j^i$". The attribute a_j^i is applied to an item x, where its information is represented by a possibility distribution $\pi_{a_j^i(x)} : D_{a_j^i} \to [0,1]$ defined on its domain $D_{a_j^i}$, which is supposed to be normalized i.e., $\exists u \in D_{a_j^i}$ such that $\pi_{a_j^i(x)}(u) = 1$. The possibility degree of p_j^i and that of its negation are computed using the possibility measure Π by $\pi(p_j^i) = \Pi(P_j^i) = \sup_{u \in P_j^i} \pi_{a_j^i(x)}(u)$ and $\pi(\neg p_j^i) = \Pi(\overline{P_j^i}) = \sup_{u \in \overline{P_j^i}} \pi_{a_j^i(x)}(u)$ respectively, where $P_j^i \subseteq D_{a_j^i}$ and $\overline{P_j^i}$ is its complement. As $\pi_{a_j^i(x)}$ is normalized, we have $\max(\pi(p_j^i), \pi(\neg p_j^i)) = 1$. The necessity degree of p_j^i is defined with the necessity measure N by $n(p_j^i) = N(P_j^i) = 1 - \pi(\neg p_j^i) = \inf_{u \in \overline{P_j^i}} (1 - \pi_{a_j^i(x)}(u))$.
The possibility degree of p_i is $\pi(p_i) = \min_{j=1}^k \pi(p_j^i)$ and that of its negation is $\pi(\neg p_i) = \max_{j=1}^k \pi(\neg p_j^i)$. These formulas $\pi(p_i)$ and $\pi(\neg p_i)$ preserve the normalization i.e., $\max(\pi(p_i), \pi(\neg p_i)) = 1$ and are respectively noted λ_i and ρ_i. The necessity degree of p_i is $n(p_i) = 1 - \pi(\neg p_i) = \min_{j=1}^k (1 - \pi(\neg p_j^i)) = \min_{j=1}^k n(p_j^i)$.
The degrees λ_i and ρ_i allow to have the following interpretations of p_i:
• $\pi(p_i) = \lambda_i$ estimates to what extent p_i is possible,
• $n(p_i) = 1 - \rho_i$ estimates to what extent p_i is certain.
The conclusion q_i of R^i is of the form "$b(x) \in Q_i$", where $Q_i \subseteq D_b$. The possibility degrees of q_i and $\neg q_i$ are respectively noted α_i and β_i. They are defined by $\begin{bmatrix} \pi(q_i) \\ \pi(\neg q_i) \end{bmatrix} = \begin{bmatrix} 1 & s_i \\ r_i & 1 \end{bmatrix} \square_{\min}^{\max} \begin{bmatrix} \lambda_i \\ \rho_i \end{bmatrix}$ where the operator $\square_{\min}^{\max}$ uses min as the product and max as the addition. The normalization $\max(\pi(p_i), \pi(\neg p_i)) = 1$ implies:

$$\alpha_i = \max(s_i, \lambda_i) \text{ and } \beta_i = \max(r_i, \rho_i).$$

The possibility distribution of the output attribute b associated to R^i is defined for any $u \in D_b$ by $\pi_{b(x)}^{*i}(u) = \begin{cases} \alpha_i \text{ if } u \in Q_i \\ \beta_i \text{ if } u \in \overline{Q_i} \end{cases}$. Finally, with n rules, the output possibility distribution is defined by a min-based conjunctive combination:

$$\pi_{b(x)}^*(u) = \min(\pi_{b(x)}^{*1}(u), \pi_{b(x)}^{*2}(u), \cdots, \pi_{b(x)}^{*n}(u)). \quad (1)$$

We introduce some additional notations. For a set output attribute value $u \in D_b$, the computation of its possibility degree is given by:

$$\pi^*_{b(x)}(u) = \min(\gamma_1, \gamma_2, \cdots, \gamma_n), \tag{2}$$

$$\text{where } \gamma_i = \pi^{*i}_{b(x)}(u) = \max(t_i, \theta_i) \text{ with } (t_i, \theta_i) = \begin{cases} (s_i, \lambda_i) & \text{if } \gamma_i = \alpha_i \\ (r_i, \rho_i) & \text{if } \gamma_i = \beta_i \end{cases}. \tag{3}$$

The relation (2) is a more convenient formulation of (1). According to (3), for each $i = 1, 2, \cdots, n$, we remark that t_i denotes a parameter (s_i or r_i) of the rule R^i and θ_i denotes either the possibility degree λ_i of the premise p_i of the rule R^i or the possibility degree ρ_i of its negation.
For a premise of a possibilistic rule, the information given by its possibility and necessity degrees can be represented by the following triplet:

Notation 1 *For a premise p, the triplet (p, sem, d) denotes either $(p, \mathsf{P}, \pi(p))$ or $(p, \mathsf{C}, n(p))$, where $sem \in \{\mathsf{P}, \mathsf{C}\}$ (P for possible, C for certain) is the semantics attached to the degree $d \in \{\pi(p), n(p)\}$.*

We introduce the following triplets according to the $\gamma_1, \gamma_2, \cdots, \gamma_n$ appearing in the relation (2). For $i = 1, 2, \cdots, n$, we set:

$$(p_i, sem_i, d_i) = \begin{cases} (p_i, \mathsf{P}, \lambda_i) & \text{if} \quad \gamma_i = \alpha_i \\ (p_i, \mathsf{C}, 1 - \rho_i) & \text{if} \quad \gamma_i = \beta_i \end{cases}. \tag{4}$$

Example 1 *Possibilistic rule-based systems have been used in medecine e.g., DIABETO [4] enables an improvement in the dietetics of diabetic patients [16]. We propose a possibilistic rule-based system for controlling the blood sugar level of a patient with type 1 diabetes (Table 1), according to some factors [3]:*

	activity (act)	*current-bloodsugar (cbs)*	*future-bloodsugar (fbs)*
R^1	*dinner, drink-coffee, lunch*	*medium, high*	*high*
R^2	*long-sleep, sport, walking*	*low, medium*	*low*
R^3	*alcohol-consumption, breakfast*	*low, medium*	*low, medium*

Table 1: rule base for the control of the blood sugar level.

The premises p_1, p_2 and p_3 of the possibilistic rules R^1, R^2 and R^3 are built using two input attributes: activity (act) and current-bloodsugar (cbs). The conclusions of the rules use the output attribute future-bloodsugar (fbs). We have D_{act} = {alcohol-consumption, breakfast, dinner, drink-coffee, long-sleep, lunch, sport, walking} and $D_{cbs} = D_{fbs}$ = {low, medium, high}. As parameters of the rules, we take $s_1 = 1$, $s_2 = 0.7$, $s_3 = 1$ and $r_1 = r_2 = r_3 = 0$. The three rules are certain [10] because we have $\pi(q_i \mid p_i) = 1$ and $r_i = \pi(\neg q_i \mid p_i) = 0$.
In our example, we assume that $\pi_{act(x)}$(drink-coffee) = 1, $\pi_{cbs(x)}$(medium) = 1 and $\pi_{cbs(x)}(low) = 0.3$, while the others elements of the domains of the input attributes have a possibility degree equal to zero. The obtained output possibility distribution is: $\langle low : 0.3, medium : 0.3, high : 1 \rangle$.

3 Justifying inference results

Farreny and Prade's approach [11] focuses on two explanatory purposes for an output attribute value $u \in D_b$, which can be formulated as two questions:

(i) How to get $\pi^*_{b(x)}(u)$ strictly greater or lower than a given $\tau \in [0,1]$?

(ii) What are the degrees of the premises justifying $\pi^*_{b(x)}(u) = \tau$?

For these two questions, the parameters of the rules s_i and r_i are set.

Regarding (i), the authors of [11] give a sufficient condition to obtain $\pi^*_{b(x)}(u) > \tau$ for a particular pair (u,τ) of their example. Taking advantage of the notations (2) and (3) we note that $\pi^*_{b(x)}(u)$ ranges between $\omega = \min(t_1, t_2, \cdots, t_n)$ and 1. Following this, a necessary and sufficient condition to obtain $\pi^*_{b(x)}(u) > \tau$ according to the degrees of premises can be easily stated:

$$\forall i \in \{j \in \{1,2,\cdots,n\} \mid t_j \leq \tau\} \text{ we have } \theta_i > \tau.$$

And similarly, $\pi^*_{b(x)}(u) < \tau$ with $\omega < \tau \leq 1$ will hold if and only if $\exists i \in \{j \in \{1,2,\cdots,n\} \mid t_j < \tau\}$ such that $\theta_i < \tau$. With these assumptions on $\theta_1, \theta_2, \cdots, \theta_n$, we can give suitable conditions on the possibility distributions of the input attributes.

Regarding (ii), [11] claim that one can directly read the possibility degrees of the premises involved in the computation of the possibility degree of an output attribute value. Their claim is sustained by a particular output attribute value u of their example. In what follows, we elaborate on this question.

3.1 Justifying the possibility degree $\pi^*_{b(x)}(u) = \tau$

We give a necessary and sufficient condition that allows us to justify $\pi^*_{b(x)}(u) = \tau$ by degrees of premises. This allows to extract the subset of premises whose degrees are involved in the computation of $\pi^*_{b(x)}(u)$. To study how the possibility degree $\pi^*_{b(x)}(u) = \tau$ with $\omega \leq \tau \leq 1$ is obtained, we introduce the following two sets J^P and J^R in order to compare the parameters $t_1, t_2, \cdots t_n$ of the rules to the degrees $\theta_1, \theta_2, \cdots, \theta_n$ of the premises in the relation (2). Intuitively, J^P (resp. J^R) collect indices where θ_i is greater (resp. lower) than t_i: in other words, the γ_i related to the rule R^i can be explained by a degree of the premise (resp. by a parameter of the rule):

$$J^P = \{i \in \{1,2,\cdots,n\} \mid t_i \leq \theta_i\} \text{ and } J^R = \{i \in \{1,2,\cdots,n\} \mid t_i \geq \theta_i\}.$$

We have $\{1,2,\cdots,n\} = J^P \cup J^R$ but J^P or J^R may be empty. We take:

$$c_\theta = \min_{i \in J^P} \theta_i \text{ and } c_t = \min_{i \in J^R} t_i \text{ (with the convention } \min_{\emptyset} = 1). \tag{5}$$

For a given output attribute value, if $J^P \neq \emptyset$ (resp. $J^R \neq \emptyset$), c_θ (resp. c_t) is the lowest possibility degree justifiable by premises (resp. by the parameters of the rules). By using the properties of the min function, we establish:

Proposition 1

$$\tau = \min(c_\theta, c_t). \tag{6}$$

When we can't explain by degrees of premises. As the degrees $\theta_1, \theta_2, \cdots, \theta_n$ of the premises are computed using the possibility distributions of the input attributes, we may have $J^P = \emptyset$. In that case, $c_\theta = 1$, $J^R = \{1, 2, \cdots, n\}$ and:

$$\pi^*_{b(x)}(u) = c_t = \min(t_1, t_2, \cdots, t_n). \tag{7}$$

Clearly, it appears that $\pi^*_{b(x)}(u)$ is independent from $\theta_1, \theta_2, \cdots, \theta_n$ and we cannot justify $\pi^*_{b(x)}(u) = \tau$ by degrees of premises. *For this reason, we suppose in the following that* $J^P \neq \emptyset$.
For a set value $u \in D_b$, we remind that the triplets (p_i, sem_i, d_i) are defined in (4) according to (2) and (3). To the non-empty set J^P we associate the following set:

$$J_{b(x)}(u) = \{(p_i, \text{sem}_i, d_i) \mid i \in J^P \text{ and } \theta_i = \tau\}. \tag{8}$$

Then, using (5), the equality (6) and the definition of $J_{b(x)}(u)$, one can check directly that we have:

Proposition 2 $J_{b(x)}(u) \neq \emptyset \Longleftrightarrow \pi^*_{b(x)}(u) = c_\theta$.

This means that when $J_{b(x)}(u) \neq \emptyset$, the set $J_{b(x)}(u)$ is formed by the premises justifying $\pi^*_{b(x)}(u) = \tau$, because if $\tau = c_\theta$, $\pi^*_{b(x)}(u)$ is the minimum of some precise degrees θ_i of premises p_i. However, if $J_{b(x)}(u) = \emptyset$, we have $\tau = c_t < c_\theta$ and then τ is the minimum of some parameters s_i or r_i. In this case, there is no way for deducing τ from $\theta_1, \theta_2, \cdots, \theta_n$ and therefore from the premises.

Example 2 *For our blood sugar level control system (section 2), according to the relation (2), we set* $\pi_{fbs(x)}(high) = \min(\gamma_1, \gamma_2, \gamma_3)$ *with* $\gamma_1 = \alpha_1 = \max(s_1, \lambda_1)$, $\gamma_2 = \beta_2 = \max(r_2, \rho_2)$ *and* $\gamma_3 = \beta_3 = \max(r_3, \rho_3)$. *We have* $\lambda_1 = \rho_2 = \rho_3 = 1$. *To apply (6), we compute* $J^P = \{1, 2, 3\}$ *and* $J^R = \{1\}$. *The possibility degree* $\tau = 1$ *of the output attribute value high can be both justified by degrees of premises* (λ_1, ρ_2 *and* ρ_3) *or a parameter of the rule* R^1. *For rules* R^2 *and* R^3, *justifications in terms of premises only can be given. As* $c_\theta = 1 = \pi_{fbs(x)}(high)$, *using (8), the following triplets are selected:*

$$J_{fbs(x)}(high) = \{(p_1, \mathsf{P}, 1), (p_2, \mathsf{C}, 0), (p_3, \mathsf{C}, 0)\}.$$

Now assume that $r_1 > 0.3$. *For* $u = low$, *(7) holds and the corresponding set* J^P *is empty: no justification in terms of premises could be given in that case.*

4 Justification and unexpectedness

In [1], a method that reduces a premise of a fuzzy if-then rule to the structure responsible for its activation degree has been defined. In this section, analogously to the fuzzy case, we define four functions $\mathscr{R}_\pi, \mathscr{R}_n, \mathscr{C}_\pi$ and $\mathscr{C}_n$ that reduce a compounded premise with respect to a threshold $\eta > 0$. The threshold η is set according to what is modelled by the possibilistic rule-base for the following purpose: if a possibility (resp. necessity) degree is higher than the threshold, it

intuitively means that the information it models is relevantly possible (resp. certain). In order to define these reduction functions for premises, we first introduce two auxiliary functions $\mathscr{P}_\pi$ and $\mathscr{P}_n$ that are defined for propositions.
Let a be an attribute with a normalized possibility distribution $\pi_{a(x)}$ on its domain D_a and a proposition p of the form "$a(x) \in P$", where $P \subseteq D_a$. We introduce the following two subsets of D_a:

$P_\pi = \{v \in P \mid \pi_{a(x)}(v) = \Pi(P)\}$ and $P_n = P \cup \{v \in \overline{P} \mid 1 - \pi_{a(x)}(v) > N(P)\}$.

The proposition related to P_π (resp. P_n) is noted p_π (resp. p_n). Let us notice that:

Proposition 3 $\overline{P_n} = \{v \in \overline{P} \mid 1 - \pi_{a(x)}(v) = N(P)\}$.

This result is a consequence of $N(P) = \inf_{v \in \overline{P}}(1 - \pi_{a(x)}(v))$. For the proposition p with its set P, $\mathscr{P}_\pi$ reduces P if $\pi(p) \geq \eta$ and $\mathscr{P}_n$ reduces $\overline{P}$ if $n(p) \geq \eta$:

$$\mathscr{P}_\pi(p) = \begin{cases} p_\pi & \text{if } \pi(p) \geq \eta \\ p & \text{if } \pi(p) < \eta \end{cases} \text{ and } \mathscr{P}_n(p) = \begin{cases} p_n & \text{if } n(p) \geq \eta \\ p & \text{if } n(p) < \eta \end{cases}.$$

We notice that $\pi(\mathscr{P}_\pi(p)) = \pi(p)$ and $n(\mathscr{P}_n(p)) = n(p)$.
In what follows, we define the four reduction functions and show how we apply them to a triplet (p, sem, d) (notation 1). We apply $\mathscr{R}_\pi$ and $\mathscr{R}_n$ to the triplets of $J_{b(x)}(u)$, see (8), to form the justification of $\pi^*_{b(x)}(u)$. Similarly, we apply $\mathscr{C}_\pi$ and $\mathscr{C}_n$ to the same triplets to extract the unexpectedness of $\pi^*_{b(x)}(u)$.

4.1 Extracting justifications: $\mathscr{R}_\pi$ and $\mathscr{R}_n$ functions

Let $p = p_1 \wedge p_2 \wedge \cdots \wedge p_k$ be a compounded premise, where p_j for $j = 1, 2, \cdots, k$, is a proposition of the form "$a_j(x) \in P_j$" with $P_j \subseteq D_{a_j}$. The function $\mathscr{R}_\pi$ (resp. $\mathscr{R}_n$) returns the structure responsible for $\pi(p)$ (resp. $n(p)$), which is the conjunction of propositions $\mathscr{P}_\pi(p_j)$ (resp. $\mathscr{P}_n(p_j)$) that make p relevantly possible (resp. certain) or not.
The reduction function $\mathscr{R}_\pi$ extends $\mathscr{P}_\pi$ in the following sense:

$$\mathscr{R}_\pi(p) = \begin{cases} \bigwedge_{j=1}^{k} \mathscr{P}_\pi(p_j) & \text{if } \pi(p) \geq \eta \\ \bigwedge_{p_j \in \{p_s \mid \pi(p_s) < \eta \text{ for } s=1,\cdots,k\}} p_j & \text{if } \pi(p) < \eta \end{cases}.$$

Similarly, the reduction function $\mathscr{R}_n$ extends $\mathscr{P}_n$ in the following sense:

$$\mathscr{R}_n(p) = \begin{cases} \bigwedge_{j=1}^{k} \mathscr{P}_n(p_j) & \text{if } n(p) \geq \eta \\ \bigwedge_{p_j \in \{p_s \mid n(p_s) < \eta \text{ for } s=1,\cdots,k\}} p_j & \text{if } n(p) < \eta \end{cases}.$$

We notice that $\pi(\mathscr{R}_\pi(p)) = \pi(p)$ and $n(\mathscr{R}_n(p)) = n(p)$.

4.2 Extracting unexpectedness: $\mathscr{C}_\pi$ and $\mathscr{C}_n$ functions

Intuitively, with respect to the threshold η, for a compounded premise $p = p_1 \wedge p_2 \wedge \cdots \wedge p_k$ that is not relevantly possible (resp. certain), $\mathscr{C}_\pi$ (resp. $\mathscr{C}_n$) returns a conjunction of propositions, called an unexpectedness, which is not involved in the determination of $\pi(p)$ (resp. $n(p)$), although relevantly possible (resp. certain).

When $\pi(p) < \eta$ and $A_p^\pi = \{p_j \mid \pi(p_j) \geq \eta \text{ for } j = 1, \cdots, k\} \neq \emptyset$, the function $\mathscr{C}_\pi$ returns the conjunction of the propositions $\mathscr{P}_\pi(p_j)$ such that $\pi(p_j) \geq \eta$:

$$\mathscr{C}_\pi(p) = \bigwedge_{p_j \in A_p^\pi} \mathscr{P}_\pi(p_j).$$

If $\pi(p) < \eta$, each proposition p_j composing p, is either used in $\mathscr{R}_\pi(p)$ or in $\mathscr{C}_\pi(p)$, according to its possibility degree $\pi(p_j)$.
Similarly, when $n(p) < \eta$ and $A_p^n = \{p_j \mid n(p_j) \geq \eta \text{ for } j = 1, \cdots, k\} \neq \emptyset$, $\mathscr{C}_n$ returns the conjunction of the propositions $\mathscr{P}_n(p_j)$ such that $n(p_j) \geq \eta$:

$$\mathscr{C}_n(p) = \bigwedge_{p_j \in A_p^n} \mathscr{P}_n(p_j).$$

If $n(p) < \eta$, each proposition p_j composing p, is either used in $\mathscr{R}_n(p)$ or in $\mathscr{C}_n(p)$, according to its necessity degree $n(p_j)$.

4.3 Justification and unexpectedness of $\pi^*_{b(x)}(u)$

To apply in an appropriate way the reduction functions $\mathscr{R}_\pi$ and $\mathscr{R}_n$ to the premise p of a triplet (p, sem, d), see notation (1), we introduce the function $\mathscr{S}_\mathscr{R}$:

$$\mathscr{S}_\mathscr{R}(p, sem, d) = \begin{cases} (\mathscr{R}_\pi(p), sem, d) \text{ if } sem = \mathsf{P} \\ (\mathscr{R}_n(p), sem, d) \text{ if } sem = \mathsf{C} \end{cases}.$$

Similarly, to apply $\mathscr{C}_\pi$ and $\mathscr{C}_n$, we introduce the function $\mathscr{S}_\mathscr{C}$:

$$\mathscr{S}_\mathscr{C}(p, sem, d) = \begin{cases} (\mathscr{C}_\pi(p), sem, \pi(\mathscr{C}_\pi(p))) \text{ if } sem = \mathsf{P}, d < \eta \text{ and } A_p^\pi \neq \emptyset \\ (\mathscr{C}_n(p), sem, n(\mathscr{C}_n(p))) \text{ if } sem = \mathsf{C}, d < \eta \text{ and } A_p^n \neq \emptyset \end{cases}.$$

The justification of $\pi^*_{b(x)}(u)$ is formed by applying $\mathscr{S}_\mathscr{R}$ to the triplets of $J_{b(x)}(u)$, see (8):

$$\text{Justification}_{b(x)}(u) = \{\mathscr{S}_\mathscr{R}(p, sem, d) \mid (p, sem, d) \in J_{b(x)}(u)\}. \tag{9}$$

The possibilistic expressions in the triplets of (9) are sufficient to justify "$b(x)$ is u at a possibility degree $\pi^*_{b(x)}(u)$". By using $\mathscr{S}_\mathscr{C}$, we obtain the unexpectedness of $\pi^*_{b(x)}(u)$ i.e., possible or certain possibilistic expressions, which may appear to be incompatible with $\pi^*_{b(x)}(u)$ while not being involved in its determination:

$$\text{Unexpectedness}_{b(x)}(u) = \{\mathscr{S}_\mathscr{C}(p, sem, d) \mid (p, sem, d) \in J_{b(x)}(u)\}. \tag{10}$$

The purpose of an unexpectedness X is to be able to formulate statements such as *"even if X, b(x) is u at a possibility degree $\pi^*_{b(x)}(u)$"*. It is in the same vein as the "even-if-because" statements studied in [6].

Example 3 *For our blood sugar level control system (section 2), we take $\eta = 0.1$ and obtain the justification of $\pi_{fbs(x)}(high)$ and its unexpectedness:*
- *Justification$_{fbs(x)}(high) = \{(\mathscr{R}_\pi(p_1), \mathsf{P}, 1), (\mathscr{R}_n(p_2), \mathsf{C}, 0), (\mathscr{R}_n(p_3), \mathsf{C}, 0)\}$.*
- *Unexpectedness$_{fbs(x)}(high) = \{(\mathscr{C}_n(p_2), \mathsf{C}, 1)\}$.*

For the premise of the rule R^1, $\mathscr{R}_\pi$ returns the conjunction of "act(x) $\in$ {drink-coffee}" and "cbs(x) $\in$ {medium}". By applying $\mathscr{R}_n$ to the premise of R^2, we obtain the proposition "act(x) $\in$ {long-sleep, sport, walking}". For the premise

of R^3, $\mathscr{R}_n$ returns "act(x) $\in$ {alcohol-consumption, breakfast}". For the premise of R^2 and that of R^3, $\mathscr{C}_n$ returns for both "cbs(x) $\in$ {low, medium}".

5 Representing explanations of possibilistic inference decisions

In this section, we represent graphically two explanations: the justification and the unexpectedness of $\pi^*_{b(x)}(u)$, (see (9) and (10)) in terms of *conceptual graphs*. The resulting conceptual graphs are visual representations of the outcomes of several analytical operations performed on the rule base that constitute explanations. Conceptual graphs are multi-graphs composed of concept nodes representing entities and relation nodes representing relationships between these entities. They were introduced by Sowa [15] and enriched by Chein and Mugnier [5]. We rely on the work of [5] for our definitions.
In the following, a *possibilistic conceptual graph* is defined as a conceptual graph where each concept node is gifted with a degree and a semantics. For each representation, we first specify its input which we call an *explanation query*. We associate an explicit explanation query to the justification of $\pi^*_{b(x)}(u)$ and another one to its unexpectedness. Each explanation query gives rise to a vocabulary, which is a simple ontology from which we define possibilistic conceptual graphs representing *statements* and a conceptual graph representing the *structure* of the explanation. One statement is called an *observed phenomenon* and represents the possibility degree $\pi^*_{b(x)}(u)$. Depending on the chosen explanation query, the other statements represent either the justification of $\pi^*_{b(x)}(u)$ or its unexpectedness. Each representation is obtained by nesting the possibilistic conceptual graphs representing the statements in the conceptual graph representing the structure.

5.1 Explanation query

To describe the vocabularies of the two explanations, we introduce the notion of *explanation query*:

Definition 1 *An explanation query is formed by a triplet $\mathcal{E} = (\mathcal{T}, b, u)$ such that:*
• *$\mathcal{T} = \{(p, sem, d)\}$ is a finite set of triplets (notation 1),*
• *b is an attribute of domain D_b with a possibility distribution $\pi^*_{b(x)} : D_b \to [0, 1]$,*
• *$u \in D_b$ is an attribute value for which the justification or the unexpectedness of its possibility degree $\pi^*_{b(x)}(u)$ is requested.*

Let us set an explanation query $\mathcal{E} = (\mathcal{T}, b, u)$, where $m = \text{card}(\mathcal{T}) \geq 1$ for which we adopt the following notations:

Notation 2 *We index the triplets of $\mathcal{T}$ as follows:*

$$\mathcal{T} = \{v^{(1)}, v^{(2)}, \cdots, v^{(m)}\} \quad ; \quad v^{(i)} = (p^{(i)}, sem^{(i)}, d^{(i)}).$$

For each triplet $v^{(i)} = (p^{(i)}, sem^{(i)}, d^{(i)}) \in \mathcal{T}$*, we set a decomposition* $p^{(i)} = p_1^{(i)} \wedge p_2^{(i)} \wedge \cdots \wedge p_{k_i}^{(i)}$ *where for each* $j = 1, 2, \cdots, k_i$ *we have:*
• $p_j^{(i)}$ *is the proposition "*$a_j^{(i)}(x) \in P_j^{(i)}$*", where* $a_j^{(i)}$ *is an attribute with a normalized possibility distribution* $\pi_{a_j^{(i)}} : D_{a_j^{(i)}} \to [0,1]$*,* $P_j^{(i)} \subseteq D_{a_j^{(i)}}$ *and* x *is an item.*
• $\mathcal{A}^{(i)} = \{a_1^{(i)}, a_2^{(i)}, \cdots, a_{k_i}^{(i)}\}$ *with card*$(\mathcal{A}^{(i)}) = k_i$*,*
• $\mathcal{S}^{(i)} = \{P_1^{(i)}, P_2^{(i)}, \cdots, P_{k_i}^{(i)}\}$ *with card*$(\mathcal{S}^{(i)}) = k_i$*.*

We take the *disjoint unions*: $\mathcal{A} = \bigcup_{1 \leq i \leq m} \mathcal{A}^{(i)}$ and $\mathcal{S} = \bigcup_{1 \leq i \leq m} \mathcal{S}^{(i)}$. These disjoint unions will allow us to define an application δ: $\mathcal{S} \to \mathcal{A}$ verifying $\delta(P_j^{(i)}) = a_j^{(i)}$ and are necessary because the domains of two distinct attributes $a_j^{(i)}$ and $a_{j'}^{(i')}$ with $i \neq i'$ may have a non-empty intersection. Therefore, the sets $P_j^{(i)}$ and $P_{j'}^{(i')}$ of the two propositions $p_j^{(i)}$ and $p_{j'}^{(i')}$ may be equal.
For our explanations, we take the following explanation queries using the justification and the unexpectedness of an output attribute value u, see (9) and (10):

$$\mathcal{E}_J = (\text{Justification}_{b(x)}(u), b, u) \quad \text{and} \quad \mathcal{E}_U = (\text{Unexpectedness}_{b(x)}(u), b, u). \tag{11}$$

5.2 Vocabulary construction

Let $\mathcal{V}^{\mathcal{E}} = (T_C, T_R, \mathcal{I}, \delta, \sigma)$ be the vocabulary associated to the explanation query $\mathcal{E} = (\mathcal{T}, b, u)$, where T_C is the set of concept types, T_R is the set of relation symbols, $\mathcal{I}$ is the set of individual markers, $\delta : \mathcal{I} \to T_C$ is an individual typing function and a relation symbol signature σ, which gives for each relation symbol of T_R the concept type of each of its arguments [5]. In $\mathcal{V}^{\mathcal{E}}$, the attribute b and the attributes in $\mathcal{A}$ are concept types. The set $\{u\}$ is an individual marker representing the attribute value u. The sets in $\mathcal{S}$ are individual markers. To any triplet $v^{(i)}$, we associate a relation symbol $inferred_{v^{(i)}}$ of arity $k_i + 1$. Therefore, a conceptual graph based on $\mathcal{V}^{\mathcal{E}}$ may contain:

- a concept node of type b and individual marker $\{u\}$,
- a concept node of type $a_j^{(i)}$ and individual marker $P_j^{(i)}$, which gives a representation of the proposition $p_j^{(i)}$,
- a relation node of type $inferred_{v^{(i)}}$, which will be linked by multi-edges to the concept node of type b and the concept nodes representing the propositions $p_1^{(i)}, p_2^{(i)}, \cdots, p_{k_i}^{(i)}$.

Additionally, for structuring the explanations, $\mathcal{V}^{\mathcal{E}}$ includes: two concept types: Phenomenon and e, a relation symbol t, and $m + 1$ individual markers that are named Statements. In a conceptual graph based on $\mathcal{V}^{\mathcal{E}}$, we may find a concept node of type Phenomenon and marker Statement_0, m concept nodes of type e and marker Statement_i for $i = 1, 2, \cdots, m$ and a relation node of type t, which will be linked by multi-edges to the $m + 1$ concept nodes that we just described.
We explicitly define $\mathcal{V}^{\mathcal{E}}$ as follows:
• $T_C = \{b\} \cup \mathcal{A} \cup \{\mathsf{e}\} \cup \{\text{Phenomenon}\}$ with $\text{card}(T_C) = 3 + \sum_{i=1}^{m} k_i$.

• $T_R = \{inferred_{v^{(i)}} | v^{(i)} \in \mathcal{T}\} \cup \{\mathsf{t}\}$ with $\text{card}(T_R) = m+1$ and such that $\text{arity}(inferred_{v^{(i)}}) = k_i + 1$ and $\text{arity}(\mathsf{t}) = m+1$.
• $\mathcal{I} = \{\{u\}\} \cup \mathcal{S} \cup \{\text{Statement}_0, \text{Statement}_1, \cdots, \text{Statement}_m\}$ with $\text{card}(\mathcal{I}) = m + 2 + \sum_{i=1}^{m} k_i$.
• $\delta : \mathcal{I} \to T_C$ such that:
  • $\{u\} \mapsto b \quad ; \quad P_j^{(i)} \mapsto a_j^{(i)}$
  • $\text{Statement}_0 \mapsto \text{Phenomenon} \quad ; \quad \text{Statement}_i \mapsto \mathsf{e}$ for $i = 1, 2, \cdots, m$.
• The signature map σ is given by:
  • $\sigma(inferred_{v^{(i)}}) = (b, a_1^{(i)}, a_2^{(i)}, \cdots, a_{k_i}^{(i)})$ for $v^{(i)} \in \mathcal{T}$
  • $\sigma(\mathsf{t}) = (\text{Phenomenon}, \mathsf{e}, \mathsf{e}, \cdots, \mathsf{e})$.

In the vocabulary $\mathcal{V}^{\mathcal{E}_J}$associated to the explanation query $\mathcal{E}_J$, see (11), the concept type e is noted "Justification" and the relation symbol t is noted "isJustifiedBy". In $\mathcal{V}^{\mathcal{E}_U}$ of $\mathcal{E}_U$, see (11), we respectively note them "Unexpectedness" and "evenIf".

5.3 Possibilistic conceptual graphs

We introduce possibilistic conceptual graphs that extend basic conceptual graphs (BG) [5] by adding two additional fields to the labels of concept nodes:

Definition 2 *A possibilistic conceptual graph (PCG) is a BG $G = (C, R, E, l)$, where C is the concept nodes set, R the relation nodes set, E is the multi-edges set and the label function l is extended by allowing a degree and a semantics in the label of any concept node $c \in C$:*

$$l(c) = (type(c) : marker(c) | sem_c, d_c)$$

The definition of a *star BG* [5] i.e., a BG restricted to a relation node and its neighbors, is naturally extended as a star PCG.

5.4 Conceptual graphs based on the vocabulary $\mathcal{V}^{\mathcal{E}}$

Given an explanation query $\mathcal{E} = (\mathcal{T}, b, u)$ (definition 1), let us specify, in a PCG $G = (C, R, E, l)$ built on the vocabulary $\mathcal{V}^{\mathcal{E}}$, the definition of the labels of the following concept nodes:
• for a concept node $c \in C$ such that $\text{type}(c) = b$ and $\text{marker}(c) = \{u\}$, we put:

$$sem_c = \mathsf{P} \text{ and } d_c = \pi^*_{b(x)}(u). \tag{12}$$

• for a concept node $c \in C$ such that $\text{type}(c) = a_j^{(i)}$ and $\text{marker}(c) = P_j^{(i)}$, we take:

$$sem_c = sem^{(i)} \text{ and } d_c = \begin{cases} \pi(p_j^{(i)}) \text{ if } sem^{(i)} = \mathsf{P} \\ n(p_j^{(i)}) \text{ if } sem^{(i)} = \mathsf{C} \end{cases}. \tag{13}$$

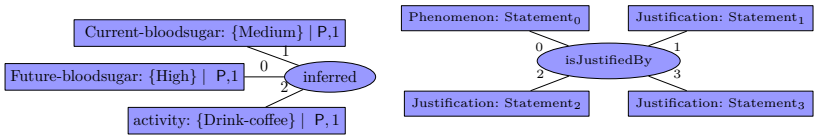
For the other concept nodes, we specify neither a degree nor a semantics.
On the vocabulary $\mathcal{V}^{\mathcal{E}}$, let us define $m+1$ PCG $D, N_1, N_2, \cdots, N_m$ and a BG R:

Definition 3 *D is defined as the PCG reduced to one concept node with label $(b : \{u\} \mid \mathsf{P}, \pi^*_{b(x)}(u))$. It is a graphical representation of a statement, which describes an observed phenomenon.*

Definition 4 *Each N_i is the star PCG where the unique relation node r_i is of type $inferred_{v^{(i)}}$ with $v^{(i)} \in \mathcal{T}$. The graph N_i contains $k_i + 1$ concept nodes: $c_b^{(i)}, c_{a_1}^{(i)}, c_{a_2}^{(i)}, \cdots, c_{a_{k_i}}^{(i)}$ of type $b, a_1^{(i)}, a_2^{(i)}, \cdots, a_{k_i}^{(i)}$ and marker $\{u\}, P_1^{(i)}, P_2^{(i)}, \cdots, P_{k_i}^{(i)}$, as in (12), (13). The multi-edges are labeled $(r_i, 0, c_b^{(i)})$ and $(r_i, j, c_{a_j}^{(i)})$ for $j = 1, 2, \cdots, k_i$. Each N_i represents graphically either a statement justifying the phenomenon represented by D or an unexpectedness statement. The link between the phenomenon and the justification statement or the unexpectedness statement is represented by a relation node of type $inferred_{v^{(i)}}$.*

Definition 5 *The graph R is the star BG where the unique relation node r is of type $\mathtt{t}$ and the $m + 1$ concept nodes are noted $c_0, c_1, \cdots, c_m$, where c_0 is of type "Phenomenon" and $c_1, c_2, \cdots, c_m$ are of type $\mathtt{e}$. Their individual markers are respectively $Statement_0, Statement_1, \cdots, Statement_m$. The multi-edges are labeled (r, j, c_j) for $j = 0, 1, \cdots, m$. R structures the explanation by representing the link between the observed phenomenon D and the statements $N_1, N_2, \cdots, N_m$.*

In Figure 1a and 1b we give examples of N_i and R respectively, with $m = 3$, $\mathtt{e}$ = "Justification" and $\mathtt{t}$ = "isJustifiedBy".



(a) Example of N_i.

(b) Example of R.

Figure 1: Examples of graphs. Nodes with a rectangular shape are concept nodes and those with an oval shape are relation nodes.

5.5 Representation of explanations

We define the representation of an explanation as a nested conceptual graph G defined by its associated tree as in [5], which is denoted $Tree(G) = (V_T, U_T, l_T)$. For our representation, the PCG $D, N_1, N_2, \cdots, N_m$ are nested in the concept nodes of R:

Definition 6 *$Tree(G) = (V_T, U_T, l_T)$ is given by:*
• *$V_T = \{R, D, N_1, N_2, \cdots, N_m\}$ is the set of nodes,*
• *$U_T = \{(R, D), (R, N_1), (R, N_2), \cdots, (R, N_m)\}$ is the set of edges and the node R is the root of $Tree(G)$,*
• *the labels of the edges are given by $l_T(R, D) = (R, c_0, D)$ and $l_T(R, N_i) = (R, c_i, N_i)$ for $i = 1, 2, \cdots, m$.*

Taking the explanation queries $\mathcal{E}_J$ and $\mathcal{E}_U$ (11), we get by definition 6, two nested conceptual graphs that represent explanations of possibilistic inference decisions.

Example 4 *We represent an explanation (Figure 2) of a decision of our blood sugar control system (section 2). It is a justification of* $\pi_{fbs(x)}(high) = 1$ *built using* $\mathcal{E}_J = (Justification_{fbs(x)}(high), fbs, high)$ *that could be in natural language: "It is possible that the patient's blood sugar level will become high. In fact, his activity is drinking coffee and his current blood sugar level is medium. In addition, it is assessed as not certain that he chose sport, walking, sleeping, eating breakfast or drinking alcohol as an activity." Its unexpectedness can also be represented.*

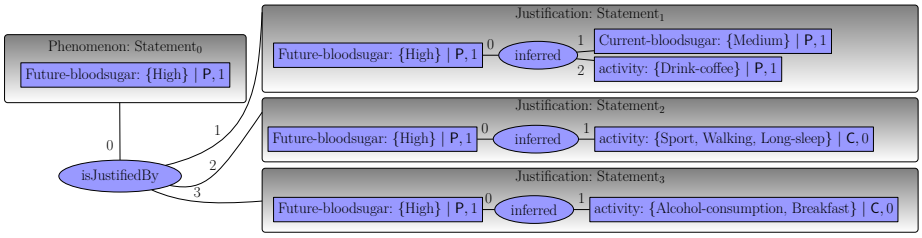


Figure 2: Representation of an explanation.

6 Conclusion

In this paper, we introduced a method to justify by a subset of rule premises the possibility degree of an output attribute value obtained by the inference of a possibilistic rule-based system. We used it to represent two kinds of explanations of possibilistic inference decisions. Natural Language Generation systems may use our representation to produce natural language explanations. Question-answering applications may also rely on our representation, as the conceptual graphs framework provides a mechanism for querying. This may lead to the development of more general extraction justification methods. Moreover, the representation of explanation may be adapted for the case of a cascade [9] and fuzzy systems.

References

[1] Baaj, I., Poli, J.P.: Natural language generation of explanations of fuzzy inference decisions. In: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). pp. 1–6. IEEE (2019)

[2] Baaj, I., Poli, J.P., Ouerdane, W.: Some insights towards a unified semantic representation of explanation for explainable artificial intelligence. In: Proceedings of the 1st Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence (NL4XAI 2019). pp. 14–19 (2019)

[3] Brown, A., Close, K.L.: Bright spots & landmines: the diabetes guide I wish someone had handed me. diaTribe Foundation (2017)

[4] Buisson, J.C., Farreny, H., Prade, H.: The development of a medical expert system and the treatment of imprecision in the framework of possibility theory. Information sciences **37**(1-3), 211–226 (1985)

[5] Chein, M., Mugnier, M.L.: Graph-based knowledge representation: computational foundations of conceptual graphs. Springer Science & Business Media (2008)

[6] Darwiche, A., Hirth, A.: On the reasons behind decisions. In: Giacomo, G.D., Catalá, A., Dilkina, B., Milano, M., Barro, S., Bugarín, A., Lang, J. (eds.) ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020). Frontiers in Artificial Intelligence and Applications, vol. 325, pp. 712–720. IOS Press (2020)

[7] Dubois, D., Prade, H.: Possibility theory: an approach to computerized processing of uncertainty. Plenum Press, New York (1988)

[8] Dubois, D., Prade, H.: Possibility theory and its applications: Where do we stand? In: Handbook of Computational Intelligence (2015)

[9] Dubois, D., Prade, H.: From possibilistic rule-based systems to machine learning-a discussion paper. In: International Conference on Scalable Uncertainty Management. pp. 35–51. Springer (2020)

[10] Farreny, H., Prade, H.: Default and inexact reasoning with possibility degrees. IEEE transactions on systems, man, and cybernetics **16**(2), 270–276 (1986)

[11] Farreny, H., Prade, H.: Positive and Negative Explanations of Uncertain Reasoning in the Framework of Possibility Theory, p. 319–333. John Wiley & Sons, Inc., USA (1992)

[12] Farreny, H., Prade, H., Wyss, E.: Approximate reasoning in a rule-based expert system using possibility theory: A case study. In: IFIP Congress. pp. 407–414 (1986)

[13] Regulation, G.D.P.: Regulation eu 2016/679 of the european parliament and of the council of 27 april 2016. Official Journal of the European Union (2016)

[14] Reiter, E., Dale, R.: Building applied natural language generation systems. Natural Language Engineering **3**(1), 57–87 (1997)

[15] Sowa, J.F.: Conceptual graphs for a data base interface. IBM Journal of Research and Development **20**(4), 336–357 (1976)

[16] Turnin, M.C.G., Beddok, R.H., Clottes, J.P., Martini, P.F., Abadie, R.G., Buisson, J.C., Dupuy, C.S., Bonneu, M., Camaré, R., Anton, J.P., et al.: Telematic expert system diabeto: new tool for diet self-monitoring for diabetic patients. Diabetes Care **15**(2), 204–212 (1992)