Explainability of possibilistic and fuzzy rule-based systems

Ismaïl Baaj

Explainability of Possibilistic and Fuzzy rule-based systems

by Ismaïl Baaj

Doctoral thesis in computer science

presented and publicly defended on January 27, 2022 at Sorbonne University

Jury :

José Maria Alonso Moral Ramón y Cajal Researcher, Universidade de Santiago de Compostela	reviewer
Sébastien Destercke CR CNRS HDR, Université de Technologie de Compiègne	reviewer
Madalina Croitoru Professeure, Université de Montpellier	examiner
Vincent Mousseau Professeur, CentraleSupélec	examiner
Patrice Perny Professeur, Sorbonne Université	examiner
Marie-Jeanne Lesot Maîtresse de Conférences HDR, Sorbonne Université	examiner
Nicolas Maudet Professeur, Sorbonne Université	director
Jean-Philippe Poli Ingénieur de recherche, CEA	advisor
Wassila Ouerdane Maîtresse de Conférences, CentraleSupélec	invited

Acknowledgements

During my thesis, I benefited from the help of many people, and it is a great pleasure to thank them.
First of all, I would especially like to express my gratitude to my main doctoral advisor Nicolas Maudet for his guidance and advice throughout my thesis. I enjoyed working with him, both from a scientific and human point of view. I am thankful to Jean-Philippe Poli for introducing me to the field of explainable AI during my master's thesis. I also thank him and Wassila Ouerdane for their advice, suggestions and comments.
I would like to thank José Maria Alonso Moral and Sébastien Destercke for carefully reviewing my entire manuscript and accepting to write a report on it. I also thank Madalina Croitoru, Vincent Mousseau, Patrice Perny and Marie-Jeanne Lesot for accepting to be examiners of my thesis. All the observations and questions of the jury members gave me many ideas for further research developments. I would also like to acknowledge the help of Marie-Jeanne Lesot and Vincent Mousseau's remarks at the mid-term of this thesis, which were very relevant.
This thesis was carried out in two laboratories, the LI3A of CEA and the LIP6 of Sorbonne University. I would like to express my gratitude to both institutions for their financial support, which gave me the opportunity to attend many conferences and seminars. Finally, I also thank Henri Prade for his numerous advices and suggestions concerning my work.
I also thank many researchers and PhD students from CEA and LIP6 who gave me good advices during my thesis: Arnaud, Régis, Vincent, Shivani, Noëlie, Hung, Etienne, Edwin, Alexis, Andrey, Sandra, Youen, Eva, Gaspard, Marvin, Yoann, Adèle and Anne-Elisabeth. I also thank my friends Theo, Florence, Odile, Oscar, Clément, Fred, Paul, Lionel, Kévin, Manon, Pomme, Emilie, Marion, Laurence, Clément, Baptiste, Quentin and Maxime for their encouragements.
Finally, I would like to highlight the role of my family. I cannot thank my parents enough for their support, they have always encouraged me to give my best and given me everything I needed to follow the directions I chose.

Contents

C Justification of the inference results of fuzzy and possibilistic rule-based systems

D Representation of explanations of fuzzy and possibilistic rule-based system inference decisions

List of Figures

Introduction

Today, advances in *Artificial Intelligence* (AI) have led to the emergence of systems capable of automating complex processes, using models that may be difficult for humans to understand [57, 70]. When humans use these AI systems, it is well known that they want to understand their behaviours and actions [66, 94], as they have more confidence in the systems that can explain their choices, assumptions and reasoning [64]. The explanatory capability of systems has become a user requirement for acceptance of their use [1], especially in human risk environments such as with autonomous vehicles or in medicine [74]. In this context, laws have recently reinforced the rights of users and allow them to claim a right of explanation in AI [1, 83]. In application of this legislation, the development of the explanatory capacity of AI systems appears today as a necessity. This requirement appeared in conjunction with the recent resurgence of interest of eXplainable Artificial Intelligence (abbreviated to XAI), a research field that aims to develop AI systems that can explain their results in a way that is understandable to humans [14]. While the first approaches for developing the explainability of AI systems date back to the 70-80s [78], this field regained popularity with the recent launch of a program by the Defense Advanced Research Projects Agency (DARPA) that attempts to bring transparency to opaque AI models such as deep neural networks [63]. Developments of principles, strategies and human-computer interaction techniques to generate effective explanations of AI systems results (see [4, 14, 73, 21]) address the emerging expectations and needs of stakeholders for AI systems [68].

In [52], Dubois, Prade and Ughetto develop the idea that information encoded on a computer may have a *negative or a positive emphasis*. Negative information acts as constraints and corresponds to statements that exclude some situations because they are impossible. Positive information models observations and corresponds to statements that describe what is possible for sure because it has been observed. These two antagonistic points of view on information allow us to distinguish different types of if-then rules to represent data and knowledge, which can be appropriately modeled in the framework of Fuzzy sets Theory and Possibility Theory.
In this thesis, based on the works [45, 50, 52, 55], we introduce explanatory

paradigms for two AI systems:

- a possibilistic rule-based system, where possibilistic rules encode negative information and
- a fuzzy rule-based system composed of possibility rules, which encode positive information.

The explanatory capacities of these systems will be developed for the following objectives:

1. the establishment of meeting points between *Knowledge Representation and Reasoning* (KRR) and *Machine Learning* (ML). They have been recently surveyed by [13].
2. the elaboration of a *processing chain* for XAI, which was proposed in [16], in order to be able to generate natural language explanations of the decisions of AI systems and to be able to evaluate the explanations.

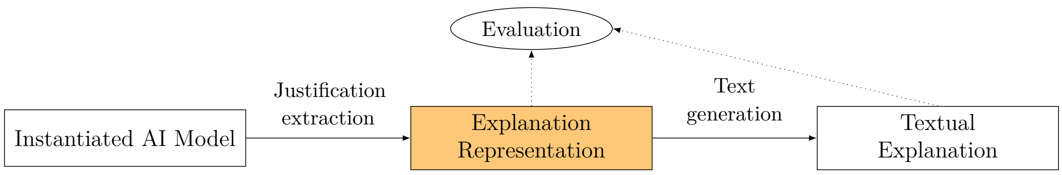


Figure 1: Proposed processing chain to generate and evaluate explanations [16]

This processing chain (Figure 1) involves at its center a *representation of an explanation* and three tasks:

- the *justification extraction* of AI system results, in order to form the content of an explanation [14, 21, 62],
- the *text generation* of an explanation, with Natural Language Generation (NLG) techniques [59, 84],
- the *evaluation* of an explanation, see [42, 75, 97, 101].

The three tasks are separated into distinct processes in order to allow a specific development of each of them and to dissociate responsibilities. In this thesis, in order to elaborate this processing chain, we focus on the task of justification extraction and the definition of a graphical representation of an explanation. From the representation, one may be able to generate natural language explanations and evaluate these explanations.

The thesis is structured as follows. In Part A, Possibility Theory and Fuzzy Set Theory are reminded. The notions of positive and negative information is

presented, as well as fuzzy rule-based systems and possibilistic rule-based systems. We also give a brief review of explanatory approaches for these systems.
In Part B, in line with our first objective of establishing links between KRR and ML, we develop a possibilistic interface between learning and if-then reasoning. Such an interface is constructed by generalizing the min-max equation system of Farreny and Prade [55], which was developed for possibilistic rule-based systems. Dubois and Prade think that this interface may allow the development of possibilistic learning methods that would be consistent with rule-based reasoning [50].
Parts C and D focus on the elaboration of the processing chain. In Part C, we introduce methods for justifying the inference results of possibilistic and fuzzy rule-based systems. Our methods allow us to form two explanations of an inference result of a rule-based system: its justification and its unexpectedness (a set of logical statements that are not involved in the determination of the considered result while being related to it). The notion of unexpectedness is inspired by that given in Simplicity Theory [40], where it aims to capture exactly what people consider surprising in a given situation [85].
In Part D, we represent these explanations in terms of conceptual graphs [34]. We also represent an explanation that is the combination of a justification and unexpectedness of an inference result. These representations let us graphically see the results of multiple analytical operations performed to generate explanations of inference decisions. From these representations, one could produce explanations in natural language, by adapting NLG systems that produce text from semantic web inputs [24, 58].

PART A

Explainable Artificial Intelligence

In this part, we give the necessary background for this thesis by reminding Possibility Theory and Fuzzy Set Theory and presenting the notions of positive and negative information (Chapter 1). This leads us also to remind possibilistic rule-based systems and fuzzy rule-based systems. We also present conceptual graphs, which is a suitable framework for representing knowledge by graphs. We then give a brief review of the explanatory approaches for rule-based systems (Chapter 2). We start with an overview of early developments for classical expert systems. We then review some approaches to develop the explanatory capabilities of fuzzy rule-based systems and possibilistic rule-based systems.

Chapter 1

Background

In this chapter, we give the necessary background for this thesis. We start by reminding Possibility Theory, which is a well-known framework for the handling of incomplete or imprecise information [43, 44] introduced by Zadeh in 1978 [108]. In our study, we focus on the possibilistic handling of rule-based systems, which was developed in the 80's [53, 56].
We then continue by presenting Fuzzy logic, which was also introduced by Zadeh [105]. It can be seen as an extension of Boolean logic, with partial truth information. We present two different types of fuzzy rules: *conjunctive fuzzy rules* and *implicative fuzzy rules* and remind how their semantics can be captured in the framework of Possibility Theory.
Finally, to represent knowledge in terms of graphs, we present the framework of conceptual graphs. In this thesis, conceptual graphs will allow us to give a graphic representation of some of our results.

1.1 Possibility Theory

Initially introduced by Zadeh and considerably developed by Dubois and Prade [43, 44], Possibility Theory is an uncertainty theory, which provides computable methods for the representation of incomplete or imprecise information. Basically, in Possibility Theory, uncertainty is modeled by two dual measures called possibility and necessity, which allows us to distinguish what is possible without being certain at all and what is certain to a some extent [50].
In the following, we give the necessary background on the possibilistic handling of rule-based systems and study the case of a cascade i.e., when a possibilistic rule-based system uses two chained sets of possibilistic rules. We also remind the necessary notions for capturing the semantics of fuzzy rules in Possibility Theory [100].

1.1.1 Possibility and necessity measures

Let U be a set. Any subset $A \subseteq U$ is called an *event*. In particular, for each $u \in U$, the singleton $\{u\}$ is called an *elementary event*. On the set 2^U, a possibility measure is defined by:

Definition 1.1 *A possibility measure on U is a map $\Pi : 2^U \to [0,1]$, which assigns a degree $\Pi(A)$ to each event $A \subseteq U$ in order to assess to what extent the event A is possible. It satisfies the following conditions:*

- $\Pi(\emptyset) = 0$ *and* $\Pi(U) = 1$,
- *For any subset* $\{A_1, A_2, \ldots, A_n\} \subseteq 2^U$, $\Pi(\bigcup_{i=1}^n A_i) = \sup_{i=1,2,\ldots,n} \Pi(A_i)$.

For any event A, if $\Pi(A)$ is equal to 1, it means that A is possible, while if $\Pi(A)$ is equal to 0, it means that A is impossible. A possibility measure Π has the following properties:

- $\Pi(A \cup \overline{A}) = \max(\Pi(A), \Pi(\overline{A})) = 1$.
- For any $A_1, A_2 \in 2^U$, if $A_1 \subseteq A_2$, then $\Pi(A_1) \leq \Pi(A_2)$. It follows that for any $A_1, A_2 \in 2^U$, we have $\Pi(A_1 \cap A_2) \leq \min(\Pi(A_1), \Pi(A_2))$.

Likewise to the notion of possibility measure, a *necessity measure* is defined by:

Definition 1.2 *A necessity measure on U is a map $N : 2^U \to [0,1]$, which assigns a degree $N(A)$ to each event $A \subseteq U$ in order to assess to what extent the event A is certain. It satisfies:*

- $N(\emptyset) = 0$ *and* $N(U) = 1$,
- *For any subset* $\{A_1, A_2, \ldots, A_n\} \subseteq 2^U$, $N(\bigcap_{i=1}^n A_i) = \inf_{i=1,2,\ldots,n} N(A_i)$.

If $N(A) = 1$, it means that A is certain. If $N(A) = 0$, the event A is not certain, but this does not mean that A is impossible. The necessity measure has the following properties:

- $N(A \cap \overline{A}) = \min(N(A), N(\overline{A})) = 0$.
- For any $A_1, A_2 \in 2^U$, if $A_1 \subseteq A_2$, then $N(A_1) \leq N(A_2)$. It follows that for any $A_1, A_2 \in 2^U$, we have $N(A_1 \cup A_2) \geq \max(N(A_1), N(A_2))$.

These two notions are dual to each other in the following sense:

- If Π is a possibility measure, then a necessity measure N is obtained by the following formula:

$$N(A) := 1 - \Pi(\overline{A}).$$

- Reciprocally, if N is a necessity measure, then a possibility measure Π is obtained by the following formula:

$$\Pi(A) := 1 - N(\overline{A}).$$

1.1.2 Possibility distribution

A possibility distribution on the set U is defined by:

Definition 1.3 *A possibility distribution π is a map $\pi : U \to [0, 1]$, which assigns to each element $u \in U$ a possibility degree $\pi(u) \in [0, 1]$. The possibility distribution is said to be* normalized *if $\exists u \in U$ such that $\pi(u) = 1$.*

Any possibility measure Π gives rise to a normalized possibility distribution π defined by the formula:

$$\pi(u) = \Pi(\{u\}), u \in U.$$

Therefore, for any subset $A \subseteq U$, we have:

$$\Pi(A) = \sup_{x \in A} \pi(x) \quad \text{and } N(A) = 1 - \Pi(\overline{A}) = \inf_{x \notin A}(1 - \pi(x)).$$

Reciprocally, a normalized possibility distribution π gives rise to a possibility measure Π defined by the formula:

$$\Pi(A) = \sup_{x \in A} \pi(x), A \subset U.$$

If a possibility distribution is associated to a variable X which takes its values on U, it is noted π_X. We have the following interpretations:

- If $\pi_X(u) = 0$ then it means that $X = u$ is a forbidden value for X.
- If $\pi_X(u) = 1$, it means that nothing prevents X from being equal to u. It is therefore *possible*, but not guaranteed possible in the common sense.

Possibility distributions were originally proposed to represent *negative information*, in the sense that they are intended (essentially) to exclude impossible elementary events [44].

Example 1.1 *Let us consider that we want to register to a master's degree course among U = { artificial_intelligence, cryptography, high_performance_computing, programming_language_theory }. After reviewing our resume, the university gives us a possibility distribution* π_{course} *that allows us to know, for each course, the possibility degree of being admitted to it:*

- $\pi_{course}(artificial\_intelligence) = 1$,
- $\pi_{course}(cryptography) = 0.5$,
- $\pi_{course}(high\_performance\_computing) = 0.3$,
- $\pi_{course}(programming\_language\_theory) = 0$.

As $\pi_{course}(artificial\_intelligence) = 1$*, the possibility distribution is normalized. The event* $A = \{cryptography, high\_performance\_computing\}$ *is assessed as possible with a degree of* $\Pi(A) = 0.5$.

Guaranteed possibility distribution

By contrast to negative information, *positive* information is represented with the help of a (*guaranteed*) possibility distribution, commonly denoted δ [46, 47]. For $u \in U$, we have the following interpretations [44]:

- $\delta_X(u) = 1$ means that the $X = u$ is really possible, i.e., it has really been observed.
- $\delta_X(u) = 0$ expresses ignorance: it means that $X = u$ has not been observed (yet: potential impossibility).

A (guaranteed) possibility distribution gives rise to a measure of guaranteed possibility Δ, which differs from Π [44]. It is defined by:

Definition 1.4 *The guaranteed possibility measure associated to the (guaranteed) possibility distribution* δ *is the map* $\Delta : 2^U \to [0, 1]$ *defined by:*

$$\text{for any } A \subseteq 2^U, \Delta(A) = \inf_{u \in A} \delta(u).$$

Δ *satisfies:*

- $\Delta(\emptyset) = 1$ *(convention),*
- *For any* $A_1, A_2 \in 2^U$*, if* $A_1 \subseteq A_2$*, then* $\Delta(A_1) \geq \Delta(A_2)$.
- *For any* $A_1, A_2 \in 2^U$*,* $\Delta(A_1 \cup A_2) = \min(\Delta(A_1), \Delta(A_2))$.

Joint possibility distribution

Let X and Y be two variables defined on U and V respectively. A joint possibility distribution $\pi_{X,Y}$ for the pair of variables (X, Y) is a map

$$\pi_{X,Y} : U \times V \rightarrow [0, 1],$$

which assigns for any pair $(u, v) \in U \times V$, a degree $\pi_{X,Y}(u, v)$ to state to what extent $X = u$ and $Y = v$ are possible values of X and Y respectively. The projections of $\pi_{X,Y}$ on each referential set are called marginal possibility distributions:

$$\forall u \in U, \pi_X(u) = \sup_{v \in V} \pi_{X,Y}(u, v) \text{ and } \forall v \in V, \pi_Y(v) = \sup_{u \in U} \pi_{X,Y}(u, v).$$

The joint possibility distribution $\pi_{X,Y}$ and its marginal possibility distributions π_X and π_Y are linked by the following relation:

$$\forall (u, v) \in U \times V, \pi_{X,Y}(u, v) \leq \min(\pi_X(u), \pi_Y(v)).$$

1.1.3 Possibilistic handling of rule-based system

A possibilistic rule-based system is composed of n if-then possibilistic rules $R^1, R^2, \ldots, R^n$. Each rule R^i has an uncertainty propagation matrix

$$\begin{bmatrix} \pi(q_i|p_i) & \pi(q_i|\neg p_i) \\ \pi(\neg q_i|p_i) & \pi(\neg q_i|\neg p_i) \end{bmatrix} = \begin{bmatrix} 1 & s_i \\ r_i & 1 \end{bmatrix},$$

which encodes the uncertainty of "if p_i then q_i" and of "if $\neg p_i$ then $\neg q_i$".
The premise p_i is of the form $p_i = p^i_1 \wedge p^i_2 \wedge \cdots \wedge p^i_k$, where each p^i_j is a proposition: "$a^i_j(x) \in P^i_j$". The attribute a^i_j is applied to an item x, where its information is represented by a possibility distribution $\pi_{a^i_j(x)} : D_{a^i_j} \rightarrow [0, 1]$ defined on its domain $D_{a^i_j}$, which is supposed to be normalized i.e., $\exists u \in D_{a^i_j}$ such that $\pi_{a^i_j(x)}(u) = 1$. The possibility degree of p^i_j and that of its negation are computed using the possibility measure Π by:

$$\pi(p^i_j) = \Pi(P^i_j) = \sup_{u \in P^i_j} \pi_{a^i_j(x)}(u) \text{ and } \pi(\neg p^i_j) = \Pi(\overline{P^i_j}) = \sup_{u \in \overline{P^i_j}} \pi_{a^i_j(x)}(u),$$

where $P^i_j \subseteq D_{a^i_j}$ and $\overline{P^i_j}$ is its complement. As $\pi_{a^i_j(x)}$ is normalized, we have $\max(\pi(p^i_j), \pi(\neg p^i_j)) = 1$. The necessity degree of p^i_j is defined with the necessity measure N by $n(p^i_j) = N(P^i_j) = 1 - \pi(\neg p^i_j) = \inf_{u \in \overline{P^i_j}}(1 - \pi_{a^i_j(x)}(u))$.
The possibility degree of p_i and that of its negation are defined by:

$$\pi(p_i) = \min_{j=1}^{k} \pi(p^i_j) \text{ and } \pi(\neg p_i) = \max_{j=1}^{k} \pi(\neg p^i_j).$$

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:

$$\begin{bmatrix} \pi(q_i) \\ \pi(\neg q_i) \end{bmatrix} = \begin{bmatrix} 1 & s_i \\ r_i & 1 \end{bmatrix} \Box_{\min}^{\max} \begin{bmatrix} \lambda_i \\ \rho_i \end{bmatrix},$$

where the operator $\Box_{\min}^{\max}$ uses min as the product and max as the addition. We still have $\max(\pi(p_i), \pi(\neg p_i)) = 1$, which implies:

$$\alpha_i = \max(s_i, \lambda_i) \text{ and } \beta_i = \max(r_i, \rho_i). \tag{1.1}$$

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

$$\pi_{b(x)}^{*}(u) = \min(\pi_{b(x)}^{*1}(u), \pi_{b(x)}^{*2}(u), \ldots, \pi_{b(x)}^{*n}(u)). \tag{1.2}$$

1.1.4 Cascade

In this case, a possibilistic rule-based system relies on $R^1, R^2, \ldots, R^n$ and a new set of m if-then possibilistic rules $R'^1, R'^2, \ldots, R'^m$, where both the conclusions of the R^i and the premises of the R'^j use the *same attribute*, establishing a chaining of the two sets of rules. In fact, each rule R'^j is of the form "if p'_j then q'_j" where p'_j is a proposition "$b(x) \in Q'_j$", Q'_j being a subset of D_b. The conclusion q'_j is of the form "$c(x) \in Q''_j$" where Q''_j is a subset of D_c, the domain of the attribute c.
The possibility degrees associated with R'^j are calculated in the same way as those of the rules R^i: $\lambda'_j = \pi(p'_j)$ and $\rho'_j = \pi(\neg p'_j)$ as p'_j is a proposition. Similarly, R'^j has an uncertainty propagation matrix with its associated parameters s'_j, r'_j.

1.2 Fuzzy Set Theory and Fuzzy logic

In the following, we remind Fuzzy Set Theory and Fuzzy logic, which were introduced by Zadeh [105]. The main goal of Fuzzy Set Theory is to represent *linguistic statements* by fuzzy sets. Based on Fuzzy Set Theory, Fuzzy logic extends Boolean logic with partial truth information and is able to deal with the vagueness of natural language.

1.2.1 Fuzzy sets

Let U be a set, which is sometimes called a reference set. The ordinary subsets of U are completely determined by their characteristic functions (also called membership functions): if $A \subset U$, the characteristic function 1_A of A is the function:

$$1_A : U \to \{0,1\} : u \mapsto 1_A(u) = \begin{cases} 1 & \text{if } u \in A \\ 0 & \text{if } u \notin A \end{cases} . \tag{1.3}$$

The notion of *fuzzy subset* of U consists in considering any function:

$$\mu : U \to [0,1] \quad : \quad \text{we replace in (1.3) the set } \{0,1\} \text{ by the interval } [0,1].$$

as a membership function of a fuzzy subset of U. In other words, to give a fuzzy subset A of U is equivalent to give its fuzzy membership function:

$$\mu_A : U \to [0,1] : u \mapsto \mu_A(u).$$

If A is a fuzzy subset of U with membership function $\mu_A : U \to [0,1]$ and B is a fuzzy subset of U with membership function $\mu_B : U \to [0,1]$, we have by definition:

$$A = B \Longleftrightarrow \forall u \in U, \mu_A(u) = \mu_B(u). \tag{1.4}$$

An *ordinary* subset $A \subset U$ is also a fuzzy subset of U with as "fuzzy" membership function its characteristic function 1_A because $\{0,1\} \subset [0,1]$.

Guaranteed possibility measure of a fuzzy set

In [49], the authors extend the definition of the guaranteed possibility measure (Definition 1.4) to the fuzzy case. Let δ_X be the (guaranteed) possibility distribution associated to the variable X on the set U. The guaranteed possibility measure of a fuzzy subset F of U, whose membership function is μ_F is defined by:

$$\Delta(F) = \inf_{u \in U} \mu_F(u) \to_g \delta_X(u), \tag{1.5}$$

where $\to_g$ is the Gödel implication defined by:

$$a \to_g b = \begin{cases} 1 & \text{if} \quad a \leq b \\ b & \text{if} \quad a > b \end{cases} .$$

Then, given a number $\alpha \in [0,1]$, the statement:

$$X \ \textit{is} \ F \ \textit{is} \ \alpha - \textit{possible},$$

means that:

$$\Delta(F) \geq \alpha.$$

1.2.2 Linguistic variables

At the root of Fuzzy Set Theory, Zadeh introduced *linguistic variables* to describe a situation, phenomena, or processes such as temperature, age, speed, etc. [106]. This notion models a variable (for example *temperature*) characterized by linguistic descriptions (e.g. *low, medium, high*):

Definition 1.5 *A linguistic variable is a triplet $a = (X, U, T_a)$ in which:*

1. *X is the name of the variable (temperature, age, speed...).*
2. *U is the domain where the variable X takes its value. The set U is called the universe of discourse.*
3. *$T_a = \{A_1, A_2, A_3, \dots\}$ is a set of fuzzy subsets of X where each fuzzy set models a linguistic description or a label of X and is called a linguistic term of the variable X.*

Let us give an example of a linguistic variable:

Example 1.2 *Let us consider a linguistic variable $a = (X, U, T_a)$ of a variable $X =$Temperature, which takes its values in $U = [0, 100]$, and has three linguistic terms $T_a = \{Low, Medium, High\}$. Its associated fuzzy partition is given in Figure 1.1.*

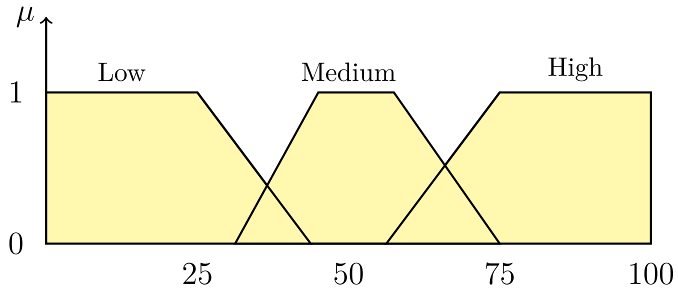


Figure 1.1: Fuzzy partition of the variable *Temperature*.

1.2.3 Fuzzy propositions

Let $a = (X, U, T_a)$ be a linguistic variable. When the available knowledge about the variable X is precise and certain, the statements that characterize X take the form $X = u$ where $u \in U$. When the available knowledge about X is imprecise or uncertain, we use a fuzzy proposition of the form "X is A", where $A \in T_a$ is a linguistic term of a.

Two different semantics can be associated to a fuzzy proposition "X is A" [100]:

- Statements of the form "X is A is *possible*" are modeled by:

$$\forall u \in U, \delta_X(u) \geq \mu_A(u),$$

where δ_X is a guaranteed possibility distribution for the variable X. Such a statement "X is A is possible" indicates that the values in A are possible to some extent. If $\delta_X(u) = 1$ then $X = u$ is a real situation, while $\delta_X(u) = 0$ means that no evidence for $X = u$ has been collected.

- Statements of the form "X *must be* in A" are modeled by:

$$\forall u \in U, \pi_X(u) \leq \mu_A(u),$$

where π_X is a possibility distribution for the variable X. Such a statement represents a constraint i.e., a fuzzy restriction on a set of possible situations: if $\mu_A(u) = 0$ then $X = u$ is an excluded value because $\pi_X(u) = 0$.

1.2.4 Fuzzy logic expressions

Fuzzy logic expressions that use the logical connectors "and", "or" and "not" can be formed from fuzzy propositions of different linguistic variables, such as "X is A and Y is B". Given two linguistic variables $a = (X, U_a, T_a)$ and $b = (Y, U_b, T_b)$ with $A \in T_a$, $i_a \in U_a$, $B \in T_b$ and $i_b \in U_b$, we compute the truth value of fuzzy logic expressions as follows:

- The truth value of "X is A and Y is B" is $T(\mu_A(i_a), \mu_B(i_b))$, where T is a t-norm.
- The truth value of "X is A or Y is B" is $S(\mu_A(i_a), \mu_B(i_b))$, where S is the t-conorm associated to T.
- The truth value of "X is not A" is $N(\mu_A(i_a))$, where N is the negation adapted to T and S.

In this thesis, we use the t-norm min and the t-conorm max, which are related by the usual negation $t :\mapsto 1-t$. From the above definitions, one can construct fuzzy rules of the form "if p then q", where p is a premise and q a conclusion, that we remind in the following.

1.2.5 Fuzzy rules

Fuzzy rules are often advocated as the basic unit of fuzzy logic-based systems [48, 100]. As we will see first in the case of a crisp rule in classical logic, a fuzzy rule is underlying *positive or negative information* [52, 67]. This basic interpretation leads to distinguish two distinct classes of fuzzy rules: conjunctive fuzzy rules and implicative fuzzy rules.

Negative and positive information

In [52], Dubois, Prade and Ughetto distinguish between the *negative* and *positive* information underlying a rule. This distinction was also revisited by [67]. A crisp rule in classical logic is of the form "if X is A then Z is O" where $A \subseteq U$ and $O \subseteq V$ are subsets of the domain of the variable X and Z respectively, links the two universes of discourse U and V by their local restrictions A and O. One can interpret such a rule from two different points of view, depending on whether one focuses on its examples or its counterexamples:

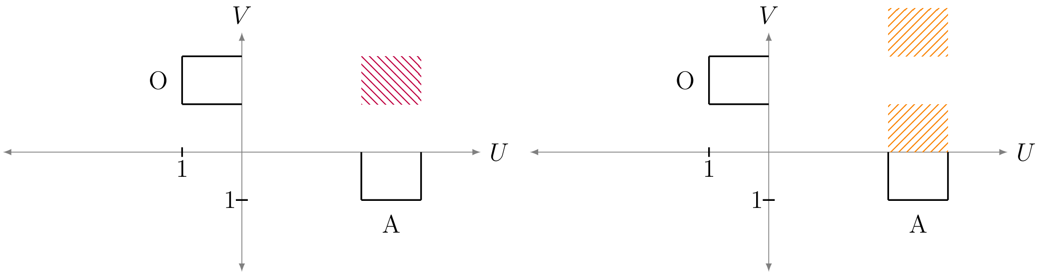
- *Positive view*: the rule is viewed as a condition of the form "if X is A then Z *can be* O" and asserts that when X takes its value in A, then *any* value in O is eligible for Z. The pairs $(u, v) \in A \times O$ form a set of examples explicitly allowed by the rule. This view is the conjunctive interpretation of the rule, emphasizing only its examples.

- *Negative view*: the rule is interpreted as a constraint of the form "if X is A then Z *must be* O" and asserts in an implicitly negative way that the values outside O are *excluded* when X takes its values in A. The pairs $(u, v) \in A \times \overline{O}$ form the set of counterexamples of the rule and are explicitly forbidden by the rule, while the pairs of the set $\overline{A \times \overline{O}}$ form the set of implicitly allowed pairs of values for (X, Z). As we have:

  $$\overline{A \times \overline{O}} = (\overline{A} \times V) \cup (U \times O) = (\overline{A} \times V) \cup (A \times O),$$

  the set $\overline{A \times \overline{O}}$ (which will be noted $\overline{A} \cup O$ in the sequel), is the disjoint union of the set of examples $A \times O$ and the set $\overline{A} \times V$ of pairs of values uncommitted by the rule. This view is the implicative interpretation of the rule, emphasizing only its counterexamples (the set $A \times \overline{O}$) and clearly corresponds to the Boolean implication in classical logic.

In conclusion, the complete representation of the rule is the pair of graphs, $(A \times O, \overline{A} \cup O)$ i.e, subsets of the cartesian product $U \times V$, made of explicitly and implicitly permitted values (u, v) (Figure 1.2). In Figure 1.2 (a), $A \times O$ represents the examples of the rule, while values outside this set are considered impossible by default [52]. In Figure 1.2 (b), $\overline{A} \cup O$ represents the counterexamples of the rule and the values outside this set are considered as possible by default [52].



(a) Examples of the rule

(b) Counterexamples of the rule

Figure 1.2: The complete representation of a crisp rule $(A \times O, \overline{A} \cup O)$.

To develop the case of fuzzy rules, we begin by giving some useful preliminaries on fuzzy graphs. Then, we review the two types of fuzzy rules in the framework of Possibility Theory.

Preliminaries on fuzzy graphs

In this thesis, a fuzzy graph F in a cartesian product $U \times V$ of classical sets is simply a fuzzy subset of the set $U \times V$. It is defined by its membership function $\mu_F : U \times V \rightarrow [0, 1]$.
In the following, two particular examples of fuzzy graphs in the set $U \times V$ will be considered: let A and O be two fuzzy subsets in U and V respectively. We have:

- The fuzzy graph $F = A \times O$ whose membership function $\mu_{A\times O}$ is defined by:

$$\forall u \in U, \forall v \in V, \mu_{A\times O}(u, v) = T(\mu_A(u), \mu_O(v)), \tag{1.6}$$

  where T is a t-norm.

- The fuzzy graph $G = \overline{A} \cup O$ whose membership function $\mu_{\overline{A}\cup O}$ is defined by:

$$\forall u \in U, \forall v \in V, \mu_{\overline{A}\cup O}(u, v) = \mu_A(u) \rightarrow \mu_O(v), \tag{1.7}$$

  where $\rightarrow$ stands for a material implication. For the general definition of a material implication, see [25, 45].

Composition of fuzzy graphs

Let F and G be fuzzy graphs in the sets $U \times V$ and $V \times W$ respectively. The *sup-min composition* of F and G is the fuzzy graph in $V \times W$, noted $F \circ G$, whose membership function $\mu_{F\circ G}$ is defined by:

$$\forall u \in U, \forall w \in W, \mu_{F\circ G}(u, w) = \sup_{v\in V} \min(\mu_F(u, v), \mu_G(v, w)).$$

The sup-min composition of fuzzy graphs is the main tool to state *the compositional rule of inference* introduced by Zadeh, see [45].
We note that if F and G are classical subsets of $U \times V$ and $V \times W$, each subset corresponds to a matrix, whose coefficients are in $[0, 1]$. Then, the sup-min composition $F \circ G$ corresponds to the max-min matrix product between these two matrices, in the same order, where the product and sum are respectively replaced by the functions min and max.
As in the matrix product case, we have a *left-sup-min composition* of a fuzzy subset A^* in U and a fuzzy graph F in $U \times V$, which is a fuzzy subset in V, noted $A^* \circ F$, whose membership function $\mu_{A^* \circ F}$ is defined by:

$$\forall v \in V, \mu_{A^* \circ F}(v) = \sup_{u \in U} \min(\mu_{A^*}(u), \mu_F(u, v)).$$

The left-sup-min composition is useful to formulate the generalized modus ponens inference pattern proposed by Zadeh and the inference of systems of parallel fuzzy if-then rules.

Fundamental dichotomy for fuzzy if-then rules

Let us consider the *if-then* rule "if X is A then Z is O" where A and O are fuzzy sets in the sets U and W respectively. By the choice of a t-norm T and a material implication $\rightarrow$, we extended the definition of the pair of graphs $(A \times O, \overline{A} \cup O)$ to the fuzzy case, see (1.6) and (1.7). This pair of fuzzy graphs $(A \times O, \overline{A} \cup O)$ in the set $U \times W$ can be understood as two modeling of the operator *then* and leads to distinguish two classes of *if-then* rules:

- A *conjunctive fuzzy rule* "if X is A then Z is O" is represented by the fuzzy graph $A \times O$. This type of rule focus on examples which are positive pieces of information, by gathering pairs of values (u, v) which are known as (more or less) feasible for (X, Z). Note that Mamdani's fuzzy inference systems [71] use conjunctive rules, where the t-norm underlying their definition is the t-norm min.

- An *implicative fuzzy rule* "if X is A then Z is O" is represented by the fuzzy graph $\overline{A} \cup O$. Implicative fuzzy rules are more natural to represent expert knowledge as they model constraints relating input and output values: they express a more or less strict constraint on the values allowed for Z, conditioned by the value taken by X. By focusing on the fuzzy graph $A \times \overline{O}$ of forbidden pairs of values, they correspond to negative pieces of information. However, as there exist numerous fuzzy extensions of the Boolean material implication, the modeling of the fuzzy graph $\overline{A} \cup O$ of implicitly allowed pairs of values (or equivalently its complement $A \times \overline{O}$), is non trivial. Note that in Zadeh's modus ponens [107], an implicative fuzzy rule plays the same role as a Boolean implication.

In what follows, we give an example of a conjunctive fuzzy rule, called *possibility rule* and an example of an implicative fuzzy rule called *certainty rule*.

Possibility rules

A conjunctive fuzzy rule "if X is A then Z is O" where the t-norm underlying to the definition of the fuzzy graph $A \times O$ is the min norm, is called a *possibility rule*. For such a conjunctive fuzzy rule, we have:

$$\forall u \in U, \forall v \in W, \mu_{A\times O}(u, v) = \min(\mu_A(u), \mu_O(v)).$$

In Possibility Theory, the operator *then* of a possibility rule "if X is A then Z is O" is modeled by a joint possibility distribution $\pi_{X,Z}$, which restricts the possible values of the pair of variables (X, Z) and satisfies [45] :

$$\forall u \in U, \forall v \in W, \min(\mu_A(u), \mu_O(v)) \leq \pi_{X,Z}(u, v). \tag{1.8}$$

Through a careful analysis [25, 45] of the relationship between the variables (X, Z), this constraint assumption (1.8) allows to justify the semantics:

"the more X is A, the more possible Z is O",

for the possibility rule "if X is A then Z is O".

Certainty rules

An implicative fuzzy rule "if X is A then Z is O" where the material implication $\rightarrow$ underlying to the definition of the fuzzy graph $\overline{A} \cup O$, is the Kleene-Dienes S-implication generated by the t-norm min [25, 45], is called a *certainty rule*. For such implicative fuzzy rule, we have:

$$\forall u \in U, \forall v \in W, \mu_{\overline{A}\cup O}(u, v) = \max(1 - \mu_A(u), \mu_O(v)).$$

In Possibility Theory, the operator *then* of a certainty rule "if X is A then Z is O" is modeled by a joint possibility distribution $\pi_{X,Z}$, which restricts the possible values of the pair of variables (X, Z) and satisfies [45] :

$$\forall u \in U, \forall v \in W, \pi_{X,Z}(u, v) \leq \max(1 - \mu_A(u), \mu_O(v)). \tag{1.9}$$

As for the possibility rules, through a careful analysis [25, 45] of the relationship between the variables (X, Z), this constraint assumption (1.9) allows to justify the semantics:

"the more X is A, the more certain Z is O",

for the certainty rule "if X is A then Z is O".
Other types of conjunctive fuzzy rules and implicative fuzzy rules have been defined, see [25].

1.2.6 Inference mechanisms

In the following, we describe the inference mechanisms for a system of parallel possibility rules and for a system of parallel certainty rules. A set of parallel rules means that the rules are of the form "If X is A_i then Z is O_i" whose inputs A_i (respectively, outputs O_i) are defined on the same universe U (respectively, V), which can be multidimensional for the premise [51].

Inference with a system of parallel possibility rules

Let $R_1, \ldots, R_n$ be a system of possibility rules. Each rule R_i is represented by a fuzzy graph $A_i \times O_i$ in $U \times V$, where A_i and O_i are fuzzy subsets in U and V respectively. The inference mechanism from a proposition "X is A^*", where A^* is a fuzzy subset in U, produces "Z is O^*" defined by the left sup-min composition $O^* = A^* \circ F$, where the membership function of the fuzzy graph F in $U \times V$ is:

$$\forall u \in U, \forall v \in V, \mu_F(u,v) = \max_i \min(\mu_{A_i}(u), \mu_{O_i}(v)).$$

This is the FATI method (FATI for *First Aggregate Then Infer*). However, as we have:

$$\begin{aligned}\forall v \in V, \mu_{O^*}(v) &= \sup_{u \in U} \min(\mu_{A^*}(u), \max_i \min(\mu_{A_i}(u), \mu_{O_i}(v)) \\ &= \max_i \sup_{u \in U} \min(\mu_{A^*}(u), \min(\mu_{A_i}(u), \mu_{O_i}(v)).\end{aligned}$$

We conclude that the output O^* is also obtained by aggregating the outputs $O_1^*, O_2^*, \ldots, O_n^*$ respectively inferred from A^* and each rule $R_1, R_2, \ldots, R_n$. This is the FITA method [100] (FITA for *First Infer Then Aggregate*), which is the Mamdani's inference method without any defuzzication (a process for obtaining a crisp value from the output aggregated fuzzy set O^*) [102].

Inference with a system of parallel certainty rules

Let $R_1, \ldots, R_n$ be a system of certainty rules. Each rule R_i is represented by a fuzzy graph $\overline{A_i} \cup O_i$ in $U \times V$, where A_i and O_i are fuzzy subsets in U and V respectively. The inference mechanism from a fact "X is A^*", where A^* is a fuzzy subset in U, produces "Z is O^*" defined by the left sup-min composition $O^* = A^* \circ G$, where the membership function of the fuzzy graph G in $U \times V$ is:

$$\forall u \in U, \forall v \in V, \mu_G(u,v) = \min_i \max(1 - \mu_{A_i}(u), \mu_{O_i}(v)).$$

This is the FATI method where $G = \cap_i(\overline{A_i} \cup O_i)$ and we have:

$$\forall v \in V, \mu_{O^*}(v) = \sup_{u \in U} \min(\mu_{A^*}(u), \min_i \max(1 - \mu_{A_i}(u), \mu_{O_i}(v)).$$

The inference by the FITA method produces "Z is O'^*" defined by aggregating n left sup-min compositions:

$$O'^* = \cap_i A^* \circ (\overline{A_i} \cup O_i),$$

where $A^* \circ (\overline{A_i} \cup O_i)$ is the left-sup-min composition of A^* and the fuzzy graph $\overline{A_i} \cup O_i$. In the case of a single rule, the FATI and the FITA methods coincide with Zadeh's generalized modus ponens.
One can check the inclusion $O^* \subseteq O'^*$ as fuzzy subsets in V. In fact, it might be rather uninformative to perform each inference $A^* \circ (\overline{A_i} \cup O_i) = O_i^*$ separately and then combine $O_1^*, O_2^*, \ldots, O_n^*$ in a conjunctive manner to get $O'^* = \cap_i O_i^*$, see [25, 45]. Therefore, for implicative fuzzy rules, only the FATI inference mechanism is used.
In the following, we study the inference of a system composed of possibility rules. For the two following examples, we apply the inference mechanism to a crisp value $u_0 \in U$, i.e. we take $A^* = \{u_0\}$. We begin with the case of one rule.

Inference with one possibility rule A possibility rule "if X is A then Z is O" is represented by the fuzzy graph $F = A \times O$ in $U \times V$, whose membership function is defined by:

$$\forall u \in U, \forall v \in V, \mu_F(u, v) = \min(\mu_A(u), \mu_O(v)).$$

The membership function of the left sup-min composition $O^* = A^* \circ F$, which defines the inferred conclusion Z is O^* is given by:

$$\begin{aligned}\forall u \in U, \forall v \in V, \mu_{O^*}(v) &= \sup_{u \in U} \min(\mu_{A^*}(u), \min(\mu_A(u), \mu_O(v)) \\ &= \min(\mu_A(u_0), \mu_O(v)).\end{aligned}$$

So, the fuzzy set O^* is the truncated fuzzy set O at the level $\mu_A(u_0)$. We continue with an example of the inference of a system composed of two possibility rules.

Inference with two possibility rules Let the two possibility rules be noted R_1, R_2. We have $O^* = A^* \circ F$, with the fuzzy graph F in $U \times V$ whose membership function is defined by:

$$\forall u \in U, \forall v \in V, \mu_F(u, v) = \max[\min(\mu_{A_1}(u), \mu_{O_1}(v)), \min(\mu_{A_2}(u), \mu_{O_2}(v))].$$

The membership function of the left sup-min composition $O^* = A^* \circ F$, which defines the inferred conclusion Z is O^* is given by:

$$\begin{aligned}\forall v \in V, \mu_{O^*}(v) &= sup_{u \in U} \min(\mu_{A^*}(u), \max_{i=1,2} \min(\mu_{A_i}(u), \mu_{O_i}(v))) \\ &= \max_{i=1,2} \sup_{u \in U} \min(\mu_{A^*}(u), \min(\mu_{A_i}(u), \mu_{O_i}(v))) \\ &= \max_{i=1,2} \min(\mu_{A_i}(u_0), \mu_{O_i}(v)).\end{aligned}$$

So, the fuzzy set O^* is the union of the truncated fuzzy set of O_1 at the level $\mu_{A_1}(u_0)$ and the truncated fuzzy set of O_2 at the level $\mu_{A_2}(u_0)$.

For the case of a possibility rule with a compounded premise, we study in the following its associated possibility distributions.

1.2.7 Possibility distributions associated to a possibility rule with a compounded premise

Let $R = (p, c)$ be a possibility rule such that:

$$p = X_1 \text{ is } A_1 \wedge X_2 \text{ is } A_2 \wedge \cdots \wedge X_k \text{ is } A_k, \tag{1.10}$$

is a conjunction of k fuzzy propositions (X_j, A_j) defined by the linguistic variables $a_j = (X_j, U_j, T_{a_j})$ and the terms $A_j \in T_{a_j}$. The conclusion c is a fuzzy proposition $c = (Z, O)$ defined by a linguistic variable $z = (Z, V, T_z)$ and $O \in T_z$.

(Guaranteed) joint possibility distribution for the input variables

The (guaranteed) joint possibility distribution $\delta_{X_1,X_2,\ldots,X_k}$ for the variables $(X_1, X_2, \ldots, X_k)$ that represent the premise p is defined on the set $U_1 \times U_2 \times \cdots \times U_k$ by (see [25]):

$$\forall (u_1, u_2, \ldots, u_k) \in U_1 \times U_2 \times \cdots \times U_k, \delta_{X_1,X_2,\ldots,X_k}(u_1, u_2, \ldots, u_k) = \min_j \mu_{A_j}(u_j). \tag{1.11}$$

From this definition, we describe as in [25, 45] the joint possibility distribution for the possibility rule $R = (p, c)$ and the possibility distribution for the conclusion $c = (Z, O)$.

Joint possibility distribution for the rule

The (guaranteed) joint possibility distribution $\delta_{X_1,X_2,\ldots,X_k;Z}$ for the possibility rule $R = (p, c)$, where p is a premise as in (1.10) and $c = (Z, O)$, is defined on the set $U_1 \times U_2 \times \cdots \times U_k \times V$ by (see [25, 45]):

$$\forall (u_1, \ldots, u_k, v) \in U_1 \times \cdots \times U_k \times V, \delta_{X_1,\ldots,X_k;Z}(u_1, \ldots, u_k, v) = \min(\min_j \mu_{A_j}(u_j), \mu_O(v)). \tag{1.12}$$

This relation (1.12) represents a *conjunctive fuzzy rule* with the choice of min as t-norm [25, 45].

Possibility distribution for the conclusion

The possibility distribution δ_Z for the variable Z of the conclusion $c = (Z, O)$ of the possibility rule $R = (p, c)$ is the marginal possibility distribution associated to the joint possibility distribution $\delta_{X_1,X_2,\ldots,X_k;Z}$ and the conclusion variable Z, see [25]:

$$\begin{aligned}\forall v \in V, \delta_Z(v) &= \sup_{(u_1,u_2,\ldots,u_k)\in U_1\times U_2\times\cdots\times U_k} \delta_{X_1,X_2,\ldots,X_k;Z}(u_1,u_2,\ldots,u_k,v) \\ &= \sup_{(u_1,u_2,\ldots,u_k)\in U_1\times U_2\times\cdots\times U_k} \min(\min_j \mu_{A_j}(u_j), \mu_O(v)). \quad (1.13)\end{aligned}$$

1.3 Conceptual graphs

In this thesis, some of our results are represented graphically. For this purpose, we use *Conceptual graphs*, which have been proposed as a knowledge representation and reasoning model mathematically founded on both logics and graph theory [18]. Conceptual graphs are multi-graphs composed of concept nodes representing entities and relation nodes representing relationships between these entities. Unlike another well-known knowledge representation formalism called RDF [65], conceptual graphs allow to represent naturally n-ary relations between entities. Conceptual graphs were introduced by Sowa [89] and enriched by Chein and Mugnier [34].
In what follows, we remind the conceptual graphs framework by studying how to build conceptual graphs and how to nest them. We begin by presenting the vocabulary used for constructing such graphs.

1.3.1 Vocabulary

A vocabulary is an ontology, which describes the different types of concepts and their relations that exist in the application domain studied [34]:

Definition 1.6 *A vocabulary $\mathcal{V} = (T_C, T_R, I, \tau, \sigma)$ is composed of two partially ordered sets T_C and T_R which are respectively the concept types and relation symbols used and a set of individual markers I. The mapping $\tau : I \rightarrow T_C$ is an individual typing function. The relation symbol signature σ gives for each relation symbol of T_R the concept type of each of its arguments.*

An individual marker is an object or an entity of the application domain. For T_C and T_R, the partial order represents a specialization. For example, if c and c' are two types of concepts with $c' < c$, it means that each instance of concept c' is also an instance of concept c.

1.3.2 Basic conceptual graph

From a vocabulary $\mathcal{V}$, one can construct a basic conceptual graph [34], which is also abbreviated BG.

Definition 1.7 *A BG is a quadruplet* $G = (C, R, E, l)$ *constructed from a vocabulary* $\mathcal{V}$*, where:*

- (C, R, E) *is a finite, undirected, bipartite multigraph, where* C *is the set of concept nodes,* R *is the set of relation nodes and* E *is the set of edges.*
- l *is a function for labeling nodes and edges which satisfies:*
  - *a concept node* $c \in C$ *is labeled by* $l(c) = (t, m)$ *where* $t = type(c) \in T_C$ *is a concept type and* $m = marker(c) \in I \cup \{*\}$ *is either the individual marker associated to the concept type* t *by the individual typing function* τ *or the generic marker* $*$ *(no individual marker). A concept node that has no individual marker is called a generic concept node.*
  - *a relation node* $r \in R$ *is labeled by* $l(r) = type(r) \in T_R$ *where* $type(r)$ *is a relation symbol of the vocabulary. The signature* σ *of* $type(r)$ *gives the types of the concept nodes to which* r *has to be linked.*
  - *The degree of a relation node* $r \in R$ *is equal to the arity of* $type(r)$*,*
  - *The edges of a relation node* $r \in R$ *are totally ordered. Typically, they are labeled from* 1 *(or* 0*) to arity(type(r)) (or arity(type(r)) - 1).*

Example 1.3 *Let us consider a vocabulary* $\mathcal{V}$*, which is an ontology to represent two humans Alice and Bob, an amount of 200 euros, and a relation borrow:*

- $T_C = \{Human, Amount\}$,
- $T_R = \{borrow\}$ *such that* $arity(borrow) = 3$,
- $I = \{Alice, Bob, 200\}$,
- $\tau : \mathcal{I} \rightarrow T_C$ *such that:*
  - *Alice* $\mapsto$ *Human,*
  - *Bob* $\mapsto$ *Human,*
  - *200* $\mapsto$ *Amount.*
- *The signature map* σ *is given by:*
  - $\sigma(borrow) = (Human, Amount, Human)$.

From the vocabulary $\mathcal{V}$, let us represent the statement "Alice borrows 200 euros to Bob" by a basic conceptual graph (Figure 1.3):

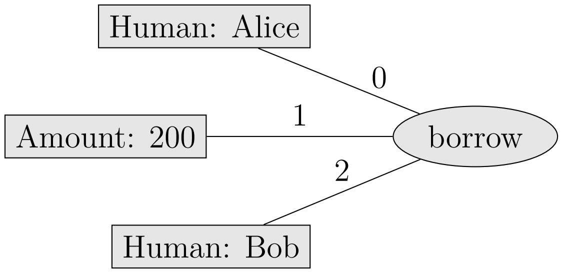


Figure 1.3: Example of basic conceptual graph. Nodes with a rectangular shape are concept nodes and those with an oval shape are relation nodes. The edge labels correspond to the numbering of the edges between a relation node and its neighbors.

A particular type of BG, called a *star BG*, has a single relation node:

Definition 1.8 *A star BG is a BG restricted to a relation node and the concept nodes that are its neighbors.*

The graph in Figure 1.3 is a star BG.

1.3.3 Nested graphs

Nested graphs are a family of conceptual graphs that have been introduced for the representation of structural information or specific contexts, zooming in and out, and distinguishing between internal and external elements of a situation. In nested graphs, a concept node can nest a concept graph.
Initially, multiple definitions of nested graphs were introduced (see [41, 61, 80, 90]). Michel Chein and Marie-Laure Mugnier proposed a definition that generalizes nested graphs as a tree of conceptual graphs [33] with an associated logical semantics [35]. This leads to the following definition of basic nested graphs [35]:

Definition 1.9 *Basic nested graphs (NBGs) are defined as follows:*

- *An elementary NBG is obtained by adding a third field to the label of each concept node c of a basic conceptual graph named $Desc(c)$ and equal to $\star\star$ (which means empty).*

- *Let G be an elementary NBG, $c_1, c_2, \ldots, c_k$ be concept nodes of G and $G_1, G_2, \ldots, G_k$ be NBGs. The graph obtained by replacing the description $\star\star$ of c_i by G_i for all $i = 1, \ldots, k$ is an NBG.*

To each NBG G, we associate a rooted tree noted $Tree(G)$ whose nodes are labeled by NBGs, and its root is noted $root(G)$ [35].

Definition 1.10 *For an NBG G, $Tree(G)$ is defined as follows:*

- *If G is an elementary NBG then $Tree(G)$ is restricted to its root, which is labeled by G.*
- *If G is the NBG obtained from the elementary NBG H, the concept nodes $c_1, c_2, \ldots, c_k$ of H and the NBGs $H_1, \ldots, H_k$, then $Tree(G)$ is constructed from $Tree(H_1), \ldots, Tree(H_k)$ by adding $root(H)$ as a root node and the edges $(root(H), c_i, root(H_i))$ for all $i = 1, \ldots, k$.*

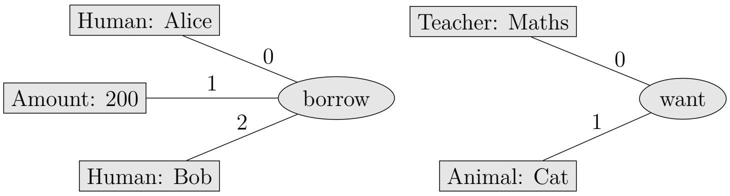
This allows us to define an NBG by a tree of BGs [34].

Definition 1.11 *A tree of BGs $T = (V_T, U_T, l_T)$ constructs an NBG G from a set of pairwise disjoint BGs $\{G_1, \ldots, G_k\}$. T is defined by:*

- *The set of nodes V_T of T is in bijection with the set of graphs $\{G_1, \ldots, G_k\}$ such that for any $i = 1, \ldots, k$, the node associated to G_i is labeled G_i.*
- *U_T is the set of arcs of T that are labeled by l_T, such that for any arc $(G_i, G_j) \in U_T$, $l_T(G_i, G_j)$ is a concept node c in G_i that can be denoted by (G_i, c, G_j).*
- *All labels are distinct, i.e. a concept node c appears at most once as an arc label.*

A logical semantics exists for NBGs, which is, for each of these graphs and as for BGs, a formula of the positive, conjunctive and existential fragment of first order logic [34].
Furthermore, there is a mapping that is called ng2bg of nested to non-nested graphs. By the existence of this mapping and their respective logical semantics, nested and non-nested graphs have the same descriptive power and are somewhat equivalent [34].

Example 1.4 *Let us consider two basic conceptual graphs G_1 and G_2 as shown in Figure 1.4:*



(a) G_1 (b) G_2 that models “A math teacher wants a cat”.

Figure 1.4: Examples of basic conceptual graphs.

We denote by c the concept node in G_1 labeled by “Human: Alice”. We form a NBG G by nesting the graph G_2 in the concept node c of G_1. The tree associated to G $Tree(G) = (V_T, U_T, l_T)$ is defined by $V_T = \{G_1, G_2\}$, $U_T = \{(G_1, G_2)\}$ and $l_T(G_1, G_2) = (G_1, c, G_2)$. It is represented in Figure 1.5.

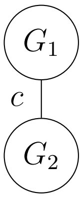


Figure 1.5: $Tree(G)$.

The resulting NBG G is represented in Figure 1.6.

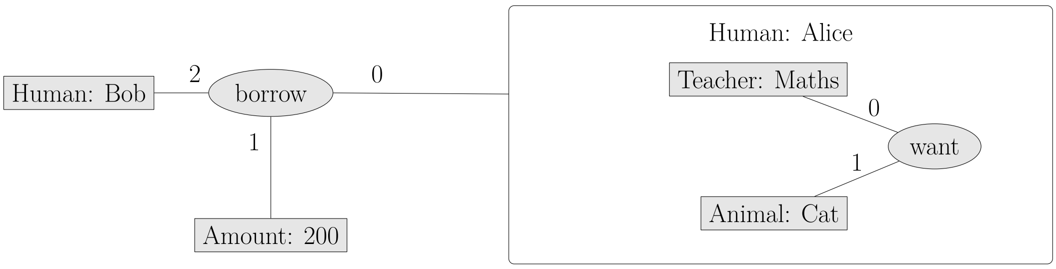


Figure 1.6: An NBG that models “Alice, a math teacher who wants a cat, borrows 200 euros to Bob”.

1.4 Conclusion

In this chapter, we reminded Possibility Theory, which is a suitable framework for the representation of imprecise and incomplete information. The possibilistic handling of rule-based system allows us to perform consistent reasoning. From a possibilistic rule-based system, we will develop in Part B an interface between learning and if-then rule-based reasoning. We will also study in Chapter 4 how we can justify the inference results of possibilistic rule-based systems.

We reminded Fuzzy set theory and Fuzzy logic. This allowed us to survey the two types of fuzzy rules, which are underlying positive and negative information and to study how to capture the semantics of fuzzy rules in the framework of Possibility Theory. We also presented two inference mechanisms of fuzzy rule-based systems: FITA and FATI. For an explanatory purpose, we study in Chapter 5 how we can justify the inference results of a fuzzy rule-based system composed of possibility rules (Mamdani rules). For such system, we investigate the semantics of its inferred conclusions.

Finally, we reminded the conceptual graph model, which allows us representing appropriately knowledge in terms of graphs. However, one can notice that it must be adapted to represent imprecise or uncertain knowledge. This will be discussed in Part D, where we represent explanations of possibilistic and fuzzy inference decisions.

Chapter 2

Explainability of rule-based systems

In this chapter, we start by giving an overview on the early developments of explainable AI systems, which date back to the 70-80s (Section 2.1). In Section 2.2, we review some approaches for developing the explanatory capabilities of fuzzy rule-based system composed of possibility rules. In Section 2.3, we focus on the min-max equation system of Farreny and Prade for developing the explanatory capabilities of a possibilistic rule-based system. Finally, we conclude (Section 2.4).

2.1 Explainability of classical expert systems

Historically, the development of the explainability of intelligent systems focused first on expert systems [78]. The explanatory stake was highlighted by the MYCIN system developed at Stanford University (an expert system allowing to identify a bacterium responsible for an infection or a blood clotting disease and to propose a treatment for it) [88]. This expert system provided a good therapy in 65% of the cases, thus outperforming all the doctors of the Stanford medical school [104]. This explanatory issue was also highlighted by Teach and Shortliffe's experiment [94], which aimed to determine physicians' expectations of computerized consultation systems. Physicians were asked to rank fifteen features that these systems could offer, and the explanatory capacity appeared to them as the most important and necessary when faced with, for example, a diagnostic error.
Early on, researchers identified two explanatory needs for AI systems: explaining how a system works and explaining how a system takes a decision [32]. These needs are sometimes referred to as *global explanation* and *local justification*, respectively [78]. During the 70-80's, the investigations lead to distinguish three types of explanations that could be generated by an expert system, in order to explain its behavior [32]:

- Explanation of the *inference results*, using the execution trace of the expert system, in which we find the sequence of rules applied in a reasoning. For instance, in MYCIN, if a user asks the system the question "Why?", the system answers by giving the user a tree structure containing the sequence of triggered rules [87]. When the user asks "How?", MYCIN allows the user to go into specific branches of the tree.

- Explanation of the (static) *knowledge base* of the expert system itself. For instance, in XPLAIN [92], authors add a knowledge base to the expert system, which contains descriptive facts about the application domain.

- Explanation of the *reasoning strategy*. In NEOMYCIN [36], it consists in the addition of meta-rules in the expert system, to explicitly represent its strategy and the relationship between rules. Such meta-rules underlie the reasoning assumptions of the system.

Following these first approaches, in the early 90s, researchers designed expert systems in order to distinguish the different types of knowledge needed to build their explanations, see [19, 81, 93]. For example, [19] separates knowledge to build an explanation into three layers: reasoning, domain knowledge and knowledge used to communicate. This kind of separation has the advantage of making the systems more modular and easier to maintain [93].

As we have seen, early work on developing the explanatory capabilities of expert systems revealed that the knowledge bases of the first expert systems were inadequate for producing explanations. Researchers stressed the importance of relying not only on the system's execution trace but also on domain knowledge, reasoning strategy, and communication knowledge.
We continue by studying approaches for the explainability of fuzzy systems.

2.2 Explainability of a fuzzy rule-based system composed of possibility rules

Zadeh is recognized as a pioneer in the field of explainable artificial intelligence [7, 26]. He introduced many paradigms that are of interest for the development of automated systems intended to interact with human agents: fuzzy sets, hedges and quantifiers, linguistic variables, approximate reasoning and fuzzy logic, computation with words, etc. His seminal work is used to build interpretable fuzzy systems [5] and is in line with the XAI challenges [26].

In what follows, we study how to explain to users the inference results of a fuzzy rule-based system composed of possibility rules that do not use a defuzzification process. To do so, we rely on the recent book by Alonso et al. [4] on the explainability of fuzzy systems. In this book, the authors study the design of interpretable fuzzy systems [3] and develop a fuzzy rule-based system offering a good compromise between interpretability/accuracy and explanatory capabilities [2]. In [3], the author states that the behaviours of a fuzzy rule based system depends on:

- its fuzzy inference process, which is set for a fuzzy rule-based system composed of a possibility rules (min norm, FITA or FATI, see Chapter 1),
- the content of its Knowledge Base (KB), which is specific to the problem domain. The KB is composed of a *database* (the linguistic variables and their related fuzzy partitions) and a *fuzzy rule base* [5].

Building the KB involves: the selection of the relevant variables for the modelling of the considered problem (for interpretability concern, the number of variables should be as small as possible [3]), the construction of the database (the modeling of linguistic variables, which implies the choice of the membership functions of its terms, the properties of the associated fuzzy partition e.g. granularity, coverage, if it is strong etc.) and the construction of the rule base (construction of the fuzzy propositions composing the fuzzy rules, from the database).
The modelling of a fuzzy system is confronted with two main objectives: accuracy and interpretability [3], which often collide. Research has been conducted to propose approaches that provide a good compromise of these two objectives, see [9]. Among them, let us mention the HILK (Highly Interpretable Linguistic Knowledge) method [6], which is an appropriate fuzzy modelling method when the interpretability of a fuzzy system is the main concern.

In 2017, authors of [12] highlighted that decisions made by fuzzy systems are better understood by users when they are explained in natural language. Then, a bibliometric analysis put forward that the fuzzy inference system seems to be a good candidate for the explainable artificial intelligence field [10]. These studies led to many approaches to elaborate the explainability of fuzzy systems:

- rLDCP [37], which is an automatic report generator for LDCP [98], a framework for Linguistic Description of Complex Phenomena that is based on the Computational Theory of Perceptions introduced by Zadeh [109]. This R package let us model complex phenomena, interpreting input data, and generating automatic reports adapted to the user needs. Given input data, a set of complex phenomena and following the LDCP methodology, it associates, to each phenomenon X, a linguistic description of the form “X is A” (for example, “the temperature is hot” for a phenomenon temperature) which is consistent with the input data. A linguistic description of a phenomenon can be deduced by Mamdani's inference.

- ExpliClas [8], which is a software that generates a textual explanation for explaining a classification made with a fuzzy rule base or a decision tree. It has been extended to support fuzzy hoeffing decision trees [11]. ExpliClas is able to generate a global explanation and a local explanation of a classification. The global explanation describes the behavior of the classifier (the list of classes of the classifier, the reliability of the classifier and an analysis of the confusion matrix). The local explanation contains information about an instance of a classification, which thus varies according to its results. If a class is inferred, the local explanation is formed by the associated fuzzy rule (or a root-to-leaf decision path for a decision tree). If a context of alternative exists with one or more other classes, the local explanation is formed by selecting for each of these classes, its name, the attribute that creates the ambiguity between these classes, and the percentage of confusion between these classes.

- For a fuzzy decision tree (and its correspondence in terms of fuzzy rules), a method for generating factual and counterfactual explanations [91]. A factual explanation is composed of feature-value pairs that together justify the root-to-leaf decision path, while a counterfactual explanation is obtained by selecting the nearest node where the root-to-leaf decision path and a path to a different class leaf diverge.

To generate natural language explanations, these approaches rely on NLG tools called surface realizers [84], e.g. SimpleNLG [60]. Such software can perform the NLG task called linguistic realization [84]: producing a syntactically, morphologically and orthographically correct text from linguistically and syntactically represented knowledge.

In the end, the approaches emphasized that the paradigms introduced by Zadeh play an important role in the construction of interpretable fuzzy systems and in the development of their explanatory capabilities [25].
Note that Possibility Theory, which is an uncertainty theory also introduced by Zadeh, also plays a role in XAI. We therefore continue by studying the explanatory capabilities of a possibilistic rule-based system in the next section.

2.3 Min-max equation system for a possibilistic rule-based system

Possibilistic rule-based systems were introduced in the 80's [53, 56]. Then, Farreny and Prade studied the problem of explaining the inference results of such systems. They considered six questions that could be asked by a user and need to be answered by explanations:

- How is obtained the output possibility distribution?
- How, mainly, is the output possibility distribution obtained? According to the authors, this question can be answered in two ways because of the term mainly. First, what are the main rules (or possibilistic propositions of the rules' premises) that determine the output possibility distribution ? Second, what are the most surprising intermediate results in the reasoning?
- How and why do such elements of the output attribute domain have a possibility degree of zero or so low?
- Which possibilistic propositions in the rule premises (if any) led to obtain a possibility degree close to zero for an output attribute value?
- How would the output possibility distribution vary depending on the possibility distribution of particular input attributes?
- Why did the system want to evaluate a particular possibilistic proposition?

From this study, these authors proposed to develop the explanatory capabilities of possibilistic rule-based systems by relying on a min-max equation system [55]. This equation system is denoted OV = MR $\blacksquare$ IV, where OV and IV are respectively named the output and the input vectors, and MR is a matrix composed of the parameters of the rules. In [50], the equation system for the case of two rules R^1 and R^2 is given:

$$\begin{bmatrix} \Pi(Q_1 \cap Q_2) \\ \Pi(Q_1 \cap \overline{Q_2}) \\ \Pi(\overline{Q_1} \cap Q_2) \\ \Pi(\overline{Q_1} \cap \overline{Q_2}) \end{bmatrix} = \begin{bmatrix} s_1 & 1 & s_2 & 1 \\ s_1 & 1 & 1 & r_2 \\ 1 & r_1 & s_2 & 1 \\ 1 & r_1 & 1 & r_2 \end{bmatrix} \Box^{\min}_{\max} \begin{bmatrix} \lambda_1 \\ \rho_1 \\ \lambda_2 \\ \rho_2 \end{bmatrix}. \tag{2.1}$$

The operator $\Box^{\min}_{\max}$ uses max as the product and min as the addition.
By performing the min-max products, the authors obtain $\Pi(Q_1 \cap Q_2) = \min(\alpha_1, \alpha_2)$, $\Pi(Q_1 \cap \overline{Q_2}) = \min(\alpha_1, \beta_2)$, $\Pi(\overline{Q_1} \cap Q_2) = \min(\beta_1, \alpha_2)$ and $\Pi(\overline{Q_1} \cap \overline{Q_2}) = \min(\beta_1, \beta_2)$. The four sets used form a partition of the output attribute domain D_b constructed from the sets Q_1, Q_2 used in the conclusions q_1, q_2 of R^1 and R^2, and their complements.

For an example of a system of three if-then possibilistic rules, the authors of [55] show that the equation system describes the output possibility distribution. They also propose to use it to perform a sensitivity analysis, depending on what is unknown (either IV or OV):

- If IV is known, they show, using their example, that they can find the possibility or necessity degrees of the rule premises, which are in IV, that justify the possibility degree of an output attribute value.
- If OV is known, they show, for a particular output attribute value u of their example and $\tau \in [0, 1]$, that we can give a sufficient condition to obtain $\pi^*_{b(x)}(u) > \tau$.

In the french version of [55], they also propose to perform sensitivity analysis by leaving some rule parameters s_i, r_i uninstantiated and observe their impact on the output possibility distribution [54].
Finally in [55], they discuss two important points. First, how we can reduce a rule premise p to return the possibilistic propositions composing the premise that together are responsible for the degree $\pi(p)$ or $n(p)$? Secondly, for a cascade, they think it would be possible to establish an input-output relationship between the equation systems associated with each set of rules of the system.

This last idea is echoed by [50]. They claim that the cascade construction would have a structural resemblance to a min-max neural network. They also put forward that such min-max equation system would be useful for developing possibilistic learning methods consistent with if-then reasoning. Such developments are in line with current research on meeting points between Knowledge Representation and Reasoning (KRR) and Machine Learning (ML) [13].

2.4 Conclusion

In this chapter, we have studied successively, the approaches to develop the explanatory capacities of classical expert systems, fuzzy rule based systems and possibilistic rule based systems. Early approaches to classical expert systems focused on building a knowledge base to construct adequate explanations for users. The explainability of fuzzy rule-based systems focus on the development of fuzzy systems that are interpretable and have explanatory capacity. We have seen that for many fuzzy rule-based systems, it is also possible to obtain explanations in natural language, using NLG tools.
Finally, we studied the min-max equation system associated with a possibilistic rule-based system. This system allows describing the output possibility distribution and to perform sensitivity analysis. The authors of [50] claim that the development of this equation system would allow to have an interface between learning and reasoning in a possibilistic setting.

PART B

Possibilistic interface between learning and reasoning

Recently, Dubois and Prade advocated the development of possibilistic learning methods that would be consistent with if-then rule-based reasoning [50]. For this purpose, the authors have proposed to represent a classification problem by the min-max equation system for a possibilistic rule-based system of Farreny and Prade [55] (reminded in Chapter 2). In particular, they highlighted the importance of the equation system for a cascade i.e., when a possibilistic rule-based system uses two sets of if-then possibilistic rules consecutively, the rules of the second set being chained with those of the first set. They suggested that the equation system associated to a cascade would have a structure somewhat similar to a min-max neural network. In this part, we give a canonical construction for the matrices governing the min-max equation system of Farreny and Prade and tackle the case of cascade. Moreover, we represent the cascade construction by a min-max neural network.

Chapter 3

Generalized min-max equation system for a possibilistic rule-based system

The work in this chapter has led to the publication of two conference papers:

- *Baaj, I., Poli, J. P., Ouerdane, W. & Maudet, N. (2021). Min-max inference for possibilistic rule-based systems. In 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.*
- *Baaj, I., Poli, J. P., Ouerdane, W. & Maudet, N. (2021). Inférence min-max pour un système à base de règles possibilistes @ In 2021 Rencontres Francophones sur la Logique Floue et ses Applications (LFA).*

In this chapter, we address a number of questions and issues raised in [50, 55] by providing an in-depth study of the min-max equation system of a possibilistic rule-based system, which we reminded in Section 2.3. This equation system describes the output possibility distribution and has been proposed to perform a sensitivity analysis [55]. In the case of n if-then possibilistic rules, we give a canonical construction of the matrices of the equation system (Section 3.1). This enables us to establish an additive formula for the output possibility distribution (Section 3.2). The output possibility distribution must be normalized for the consistency of the rules and to deal with a cascade. Using the additive formula and the equation system, we give a necessary and sufficient condition for the output possibility distribution to be normalized (Section 3.2). We also determine minimal input solutions for the normalization, when it is possible.
In Section 3.3, we calculate explicitly the measures of possibility and necessity of any subset of the output attribute domain. All these works allow us to deal with a cascade (Section 3.4) and we associate to the cascade construction a min-max neural network that describes it. We end by giving a concrete example, and some perspectives (Section 3.5).

Before introducing our work, we give some notations. In this chapter, all the matrices have their coefficients in $[0,1]$. To any matrix $A = [a_{ij}]$, we associate the matrix $A^\circ = [1 - a_{ij}]$. We have the following property: $(A^\circ)^\circ = A$.
Let A and B be matrices of respective size (n, m) and (m, p). The transformation $A \mapsto A^\circ$ switches the two matrix products in the following sense:

$$(A \Box_{\max}^{\min} B)^\circ = A^\circ \Box_{\min}^{\max} B^\circ \text{ and } (A \Box_{\min}^{\max} B)^\circ = A^\circ \Box_{\max}^{\min} B^\circ.$$

$\Box_{\max}^{\min}$ (resp. $\Box_{\min}^{\max}$) is the matricial product where we take min as addition and max as product (resp. max as addition and min as product).
Finally, we introduce an operator denoted $\boxdot_{\min}$ that takes the mininum of the coefficients of each row in a matrix.

3.1 Generalized equation system

In this section, we use a possibilistic rule-based system with a set of n if-then possibilistic rules $R^1, R^2, \ldots, R^n$. We introduce the generalized equation system, which we note:

$$O_n = M_n \Box_{\max}^{\min} I_n.$$

For $n = 2$, our construction is equivalent to the construction previously recalled, see (2.1). To understand the output vector O_n of the equation system, we introduce an explicit partition of the output attribute domain D_b. Moreover, this partition is directly linked to a matrix B_n (subsection 3.1.4) that we construct inductively with respect to the number of rules.

3.1.1 Partition and settings

From the sets $Q_1, Q_2, \ldots, Q_n$ used in the conclusions of the rules and their complements, for each $i = 1, 2, \ldots, n$, we define $(E_k^{(i)})_{1 \leq k \leq 2^i}$ a partition of D_b by the following two conditions:

- $E_1^{(1)} = Q_1$ and $E_2^{(1)} = \overline{Q_1}$ (3.1a)

and for $i > 1$:

- $$E_k^{(i)} = \begin{cases} E_k^{(i-1)} \cap Q_i & \text{if } 1 \leq k \leq 2^{i-1} \\ E_{k-2^{i-1}}^{(i-1)} \cap \overline{Q_i} & \text{if } 2^{i-1} < k \leq 2^i \end{cases}. \quad (3.1b)$$

For $i = 1, 2, \ldots, n$, we define matrices M_i, I_i and B_i according to:

- the sequences $s_1, s_2, \ldots, s_i$ and $r_1, r_2, \ldots, r_i$ for M_i,
- the sequences $\lambda_1, \lambda_2, \ldots, \lambda_i$ and $\rho_1, \rho_2, \ldots, \rho_i$ for I_i,
- the sequences $\alpha_1, \alpha_2, \ldots, \alpha_i$ and $\beta_1, \beta_2, \ldots, \beta_i$ for B_i.

In the following, we construct the matrices I_i, M_i and B_i.

3.1.2 Construction of M_i

For $i = 1$, we take $M_1 = \begin{bmatrix} s_1 & 1 \\ 1 & r_1 \end{bmatrix}$. For $i > 1$, we define M_i of size $(2^i, 2i)$ by the following block matrix construction:

$$M_i = \left[\begin{array}{c|c} M_{i-1} & S_i \\ \hline M_{i-1} & R_i \end{array}\right],$$

where $S_i = \begin{bmatrix} s_i & 1 \\ s_i & 1 \\ \vdots & \vdots \\ s_i & 1 \end{bmatrix}$ and $R_i = \begin{bmatrix} 1 & r_i \\ 1 & r_i \\ \vdots & \vdots \\ 1 & r_i \end{bmatrix}$ are of size $(2^{i-1}, 2)$.

We note $N_1, N_2, \ldots, N_{2^i}$ the rows of M_i.

3.1.3 Construction of I_i

For $i = 1$, we take $I_1 = \begin{bmatrix} \lambda_1 \\ \rho_1 \end{bmatrix}$. For $i > 1$, we define I_i of size $(2i, 1)$:

$$I_i = \left[\begin{array}{c} I_{i-1} \\ \hline \lambda_i \\ \rho_i \end{array}\right] = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{2i} \end{bmatrix},$$

where $\theta_{2j-1} = \lambda_j$ and $\theta_{2j} = \rho_j$ for $j = 1, 2, \ldots, i$.

3.1.4 Construction of B_i

For $i = 1$, we take $B_1 = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix}$. For $i > 1$, we define B_i of size $(2^i, i)$ by the following block matrix construction with B_{i-1}:

$$B_i = \left[\begin{array}{c|c} & \alpha_i \\ B_{i-1} & \alpha_i \\ & \vdots \\ & \alpha_i \\ \hline & \beta_i \\ B_{i-1} & \beta_i \\ & \vdots \\ & \beta_i \end{array}\right].$$

We note that the rows $L_1, L_2, \ldots, L_{2^i}$ of B_i are related to the rows $L'_1, L'_2, \ldots, L'_{2^{i-1}}$ of B_{i-1} by the following result:

$$L_k = \begin{cases} (L'_k, \alpha_i) & \text{if} \quad 1 \leq k \leq 2^{i-1} \\ (L'_{k-2^{i-1}}, \beta_i) & \text{if} \quad 2^{i-1} < k \leq 2^i \end{cases}. \tag{3.2}$$

3.1.5 Relations between B_i and M_i

We can recover the matrix M_i from the matrix B_i in an explicit way and vice versa.
Remind that the size of B_i and the size of M_i are respectively equal to $(2^i, i)$ and $(2^i, 2i)$. Let k be an integer in $\{1, 2, \ldots, 2^i\}$. Denote by $N_k = \begin{bmatrix} t_1 & \cdots & t_j & \cdots & t_{2i-1} & t_{2i} \end{bmatrix}$ with $t_j \in \{1, r_*, s_*\}$ and $L_k = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_i \end{bmatrix}$ with $\gamma \in \{\alpha, \beta\}$, the corresponding rows of M_i and B_i respectively.
For any $j = 1, 2, \ldots, i$, we have:

$$(t_{2j-1}, t_{2j}) = \begin{cases} (s_j, 1) \text{ if } \gamma_j = \alpha_j \\ (1, r_j) \text{ if } \gamma_j = \beta_j \end{cases} \text{ and } \gamma_j = \begin{cases} \alpha_j \text{ if } (t_{2j-1}, t_{2j}) = (s_j, 1) \\ \beta_j \text{ if } (t_{2j-1}, t_{2j}) = (1, r_j) \end{cases}.$$

In this reasoning, we ignore the particular expression that the parameters α_j, β_j can have: $\alpha_j = \max(s_j, \lambda_j)$ and $\beta_j = \max(r_j, \rho_j)$. All the above claims can be easily proved by recurrence on $i = 1, 2, \ldots$.

3.1.6 Relation between B_i and the partition

For $k \in \{1, 2, \ldots, 2^i\}$, let $L_k = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_i \end{bmatrix}$ be any row of the matrix B_i with $\gamma \in \{\alpha, \beta\}$. Then the corresponding set $E_k^{(i)}$ of the partition is equal to:

$$E_k^{(i)} = T_1 \cap T_2 \cdots \cap T_i \text{ with } T_j = \begin{cases} Q_j & \text{if } \gamma_j = \alpha_j \\ \overline{Q_j} & \text{if } \gamma_j = \beta_j \end{cases}. \tag{3.3}$$

As it is clear for $i = 1$, this result is deduced from the description of the rows of B_i by the rows of B_{i-1}, see (3.2).

3.1.7 Coefficients of $\boxdot_{\min} B_i$

For any $i = 1, 2, \ldots, n$, we set:

$$\boxdot_{\min} B_i = [o_k^{(i)}]_{1 \le k \le 2^i}.$$

For any $k \in \{1, 2, \ldots, 2^i\}$, we can deduce the following relations between the coefficients of $\boxdot_{\min} B_i$ and those of $\boxdot_{\min} B_{i-1}$:

$$o_k^{(i)} = \begin{cases} \min(o_k^{(i-1)}, \alpha_i) & \text{if } 1 \le k \le 2^{i-1} \\ \min(o_{k-2^{i-1}}^{(i-1)}, \beta_i) & \text{if } 2^{i-1} < k \le 2^i \end{cases}. \tag{3.4}$$

This result is directly deduced from (3.2) and the associativity of the min function. Finally, we obtain:

Theorem 3.1 *The min-max matrix product of M_i by the column-matrix I_i is obtained by applying the operator $\boxdot_{min}$ to the matrix B_i:*

$$M_i \Box_{\max}^{\min} I_i = \boxdot_{\min} B_i. \tag{3.5}$$

We prove (3.5) inductively:

Proof 3.1 *For $i = 1$, $M_1 \Box_{\max}^{\min} I_1 = \begin{bmatrix}\alpha_1 \\ \beta_1\end{bmatrix} = \boxdot_{\min} B_1$.*
Suppose that the relation (3.5) is true up to $i-1$. We will show it for i. So, for $k = 1, 2, \ldots, 2^{i-1}$ we have:

$$o_k^{(i-1)} = N'_k \Box_{\max}^{\min} I_{i-1}.$$

where N'_k is the k-th row of M_{i-1}.

For any $k = 1, 2, \ldots, 2^i$, let us prove that the coefficient $o_k^{(i)}$ of $\boxdot_{\min} B_i$ is equal to $N_k \Box_{\max}^{\min} I_i$, N_k being the k-th row of M_i:

- *if $1 \le k \le 2^{i-1}$:*

$$\begin{aligned} o_k^{(i)} &= \min(o_k^{(i-1)}, \alpha_i) \\ &= \min(N'_k \Box_{\max}^{\min} I_{i-1}, \max(s_i, \lambda_i)) \\ &= \min(N'_k \Box_{\max}^{\min} I_{i-1}, \max(s_i, \lambda_i), \max(1, \rho_i)) \\ &= N_k \Box_{\max}^{\min} I_i. \end{aligned}$$

- *if $2^{i-1} < k \le 2^i$:*

$$\begin{aligned} o_k^{(i)} &= \min(o_{k-2^{i-1}}^{(i-1)}, \beta_i) \\ &= \min(N'_{k-2^{i-1}} \Box_{\max}^{\min} I_{i-1}, \max(r_i, \rho_i)) \\ &= \min(N'_{k-2^{i-1}} \Box_{\max}^{\min} I_{i-1}, \max(1, \lambda_i), \max(r_i, \rho_i)) \\ &= N_k \Box_{\max}^{\min} I_i. \end{aligned}$$

This proves coefficient by coefficient:

$$O_i = \boxdot_{\min} B_i,$$

which is exactly (3.5).

Let us illustrate the equation system by an example.

Example 3.1 *We consider a possibilistic rule-based system composed of $n = 3$ rules. The sets of the partition $(E_k^{(3)})_{1 \le k \le 8}$ are the following: $Q_1 \cap Q_2 \cap Q_3$,*

$\overline{Q}_1 \cap Q_2 \cap Q_3$, $Q_1 \cap \overline{Q}_2 \cap Q_3$, $\overline{Q}_1 \cap \overline{Q}_2 \cap Q_3$, $Q_1 \cap Q_2 \cap \overline{Q}_3$, $\overline{Q}_1 \cap Q_2 \cap \overline{Q}_3$, $Q_1 \cap \overline{Q}_2 \cap \overline{Q}_3$ and $\overline{Q}_1 \cap \overline{Q}_2 \cap \overline{Q}_3$. We check (3.5) by direct calculation:

$$
\begin{aligned}
O_3 &= M_3 \square^{\min}_{\max} I_3 \\
&= \begin{bmatrix} s_1 & 1 & s_2 & 1 & s_3 & 1 \\ 1 & r_1 & s_2 & 1 & s_3 & 1 \\ s_1 & 1 & 1 & r_2 & s_3 & 1 \\ 1 & r_1 & 1 & r_2 & s_3 & 1 \\ s_1 & 1 & s_2 & 1 & 1 & r_3 \\ 1 & r_1 & s_2 & 1 & 1 & r_3 \\ s_1 & 1 & 1 & r_2 & 1 & r_3 \\ 1 & r_1 & 1 & r_2 & 1 & r_3 \end{bmatrix} \square^{\min}_{\max} \begin{bmatrix} \lambda_1 \\ \rho_1 \\ \lambda_2 \\ \rho_2 \\ \lambda_3 \\ \rho_3 \end{bmatrix} \\
&= \begin{bmatrix} \min(\alpha_1, \alpha_2, \alpha_3) \\ \min(\beta_1, \alpha_2, \alpha_3) \\ \min(\alpha_1, \beta_2, \alpha_3) \\ \min(\beta_1, \beta_2, \alpha_3) \\ \min(\alpha_1, \alpha_2, \beta_3) \\ \min(\beta_1, \alpha_2, \beta_3) \\ \min(\alpha_1, \beta_2, \beta_3) \\ \min(\beta_1, \beta_2, \beta_3) \end{bmatrix} \\
&= \boxdot_{\min} B_3.
\end{aligned}
$$

3.2 Equation system properties

In this section, we study the properties of the equation system by first establishing an additive formula for $\pi^*_{b(x)}$ from the partition $(E_k^{(i)})_{1 \le k \le 2^i}$. We give a necessary and sufficient condition for the normalization of $\pi^*_{b(x)}$. Then we show that, by deleting the empty sets of the partition and the corresponding rows of O_i, M_i and B_i, we get matrices $\mathcal{O}_i$, $\mathcal{M}_i$ and $\mathcal{B}_i$ with a reasonable number of rows. We also study the solutions for the normalization and how to rebuild the equation system if we remove a rule.

3.2.1 Additive formula for $\pi^*_{b(x)}$

We use the coefficients of $O_i = \boxdot_{\min} B_i$, see (3.5), and introduce the characteristic functions $\mu_{E_1^{(i)}}, \mu_{E_2^{(i)}}, \ldots, \mu_{E_{2^i}^{(i)}}$ of the sets $E_1^{(i)}, E_2^{(i)}, \ldots, E_{2^i}^{(i)}$.

Theorem 3.2 *The output possibility distribution* $\pi^*_{b(x),i}$ *associated to the first i rules is:*

$$\pi^*_{b(x),i} = \sum_{1\leq k\leq 2^i} o_k^{(i)} \mu_{E_k^{(i)}}. \tag{3.6}$$

We prove (3.6) inductively:

Proof 3.2 *For* $i = 1$, *we have* $O_1 = \begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix}$, $\mu_{E_1^{(1)}} = \mu_{Q_1}$ *and* $\mu_{E_2^{(1)}} = \mu_{\overline{Q_1}}$. *The sum* $\alpha_1\mu_{Q_1} + \beta_1\mu_{\overline{Q_1}}$ *is equal to* $\pi^*_{b(x),1}$. *Suppose that the relation (3.6) is true up to* $i-1$ *and show it for* i. *Let* $u \in D_b$, *we must show, in the case where we have i rules:*

$$\pi^*_{b(x),i}(u) = \begin{cases} \min(\pi^*_{b(x),i-1}(u), \alpha_i) & \text{if } u \in Q_i \\ \min(\pi^*_{b(x),i-1}(u), \beta_i) & \text{if } u \in \overline{Q_i} \end{cases} = \sum_{1\leq k\leq 2^i} o_k^{(i)} \mu_{E_k^{(i)}}(u).$$

We remind that the sets $(E_k^{(i)})_{1\leq k\leq 2^i}$ *form a partition of the set* D_b. *Therefore, there exists an unique index* k_0 *such that* $u \in E_{k_0}^{(i)}$. *This means that we can calculate* $\mu_{E_k^{(i)}}(u)$ *for all* $1 \leq k \leq 2^i$ *such that:*

$$\mu_{E_k^{(i)}}(u) = \begin{cases} 1 & \text{if} \quad k = k_0 \\ 0 & \text{if} \quad k \neq k_0 \end{cases}.$$

So we can rewrite:

$$\sum_{1\leq k\leq 2^i} o_k^{(i)} \mu_{E_k^{(i)}}(u) = o_{k_0}^{(i)} = \begin{cases} \min(o_{k_0}^{(i-1)}, \alpha_i) & \text{if } u \in Q_i \\ \min(o_{k_0-2^{i-1}}^{(i-1)}, \beta_i) & \text{if } u \in \overline{Q_i} \end{cases}.$$

To see that we have:

$$\pi^*_{b(x),i}(u) = o_{k_0}^{(i)},$$

we must distinguish two cases: if $u \in Q_i$ *and if* $u \in \overline{Q_i}$.

- *If* $u \in Q_i$, *we have* $1 \leq k_0 \leq 2^{i-1}$, $E_{k_0}^{(i)} = E_{k_0}^{(i-1)} \cap Q_i$ *and* $o_{k_0}^{(i)} = \min(o_{k_0}^{(i-1)}, \alpha_i)$.

By the recurrence hypothesis, we have with $i-1$ *rules:*

$$\pi^*_{b(x),i-1}(u) = \sum_{1\leq k\leq 2^{i-1}} o_k^{(i-1)} \mu_{E_k^{(i-1)}}(u) = o_{k_0}^{(i-1)}.$$

Finally, by associativity of the min *function:*

$$\begin{aligned}\pi^*_{b(x),i}(u) &= \min(\pi^{*1}_{b(x)}(u), \pi^{*2}_{b(x)}(u), \dots, \pi^{*i}_{b(x)}(u)) \\ &= \min(\min(\pi^{*1}_{b(x)}(u), \pi^{*2}_{b(x)}(u), \dots, \pi^{*i-1}_{b(x)}(u)), \pi^{*i}_{b(x)}(u)) \\ &= \min(\pi^*_{b(x),i-1}(u), \alpha_i) \\ &= \min(o^{(i-1)}_{k_0}, \alpha_i) \\ &= o^{(i)}_{k_0}.\end{aligned}$$

• *If* $u \in \overline{Q_i}$*, we have* $2^{i-1} < k_0 \leq 2^i$*,* $E^{(i)}_{k_0} = E^{(i-1)}_{k_0-2^{i-1}} \cap \overline{Q_i}$ *and* $o^{(i)}_{k_0} = \min(o^{(i-1)}_{k_0-2^{i-1}}, \beta_i)$*.*
As $u \in E^{(i-1)}_{k_0-2^{i-1}}$*, we have for all* $1 \leq k \leq 2^{i-1}$*:*

$$\mu_{E^{(i-1)}_k}(u) = \begin{cases} 1 & \text{if} \quad k = k_0 - 2^{i-1} \\ 0 & \text{if} \quad k \neq k_0 - 2^{i-1} \end{cases}.$$

By the recurrence hypothesis, we have with $i-1$ *rules:*

$$\pi_{b(x),i-1}(u) = \sum_{1 \leq k \leq 2^{i-1}} o^{(i-1)}_{k,j} \, \mu_{E^{(i-1)}_k}(u) = o^{(i-1)}_{k_0-2^{i-1}}.$$

Finally, by associativity of the min *function:*

$$\begin{aligned}\pi^*_{b(x),i}(u) &= \min(\pi^{*1}_{b(x)}(u), \pi^{*2}_{b(x)}(u), \dots, \pi^{*i}_{b(x)}(u)) \\ &= \min(\min(\pi^{*1}_{b(x)}(u), \pi^{*2}_{b(x)}(u), \dots, \pi^{*i-1}_{b(x)}(u)), \pi^{*i}_{b(x)}(u)) \\ &= \min(\pi^*_{b(x),i-1}(u), \beta_i) \\ &= \min(o^{(i-1)}_{k_0-2^{i-1}}, \beta_i) \\ &= o^{(i)}_{k_0}.\end{aligned}$$

As a consequence, $\forall u \in D_b$, there is a unique index k_0 such that $u \in E^{(i)}_{k_0}$ and $\pi^*_{b(x),i}(u) = o^{(i)}_{k_0}$. From this, we deduce that $\pi^*_{b(x),i}$ is normalized if and only if:

$$\exists k \in \{1, 2, \dots, 2^i\} \text{ such that } E^{(i)}_k \neq \emptyset \text{ and } o^{(i)}_k = 1. \tag{3.7}$$

Moreover, from (3.6), we deduce that the possibility measure of each *non-empty* set $E^{(i)}_k$ of the partition is equal to $o^{(i)}_k$. It is then natural to introduce:

$$J = \left\{ k \in \{1, 2, \dots, 2^i\} \;\middle|\; E^{(i)}_k \neq \emptyset \right\} \text{ and } \omega = \text{card}(J).$$

Considering $\text{card}(D_b) = d$, we have $\omega \leq \min(d, 2^i)$. We may arrange the elements of J as a strictly increasing sequence: $1 \leq k_1 < k_2 < \dots < k_\omega \leq 2^i$. We have:

$$[\Pi(E^{(i)}_k)]_{k \in J} = [o^{(i)}_k]_{k \in J}.$$

Thus, in what follows, we note $\mathcal{O}_i$, $\mathcal{M}_i$ and $\mathcal{B}_i$, the matrices obtained from O_i, M_i and B_i respectively, by deleting each row whose index is not in J.

Let us illustrate (3.6) with the following example:

Example 3.2 *We continue with a possibilistic rule-based system composed of $n = 3$ rules. The characteristic functions of the partition $(E_k^{(3)})_{1\leq k\leq 8}$ are $\mu_{Q_1\cap Q_2\cap Q_3}$, $\mu_{\overline{Q_1}\cap Q_2\cap Q_3}$, $\mu_{Q_1\cap\overline{Q_2}\cap Q_3}$, $\mu_{\overline{Q_1}\cap\overline{Q_2}\cap Q_3}$, $\mu_{Q_1\cap Q_2\cap\overline{Q_3}}$, $\mu_{\overline{Q_1}\cap Q_2\cap\overline{Q_3}}$, $\mu_{Q_1\cap\overline{Q_2}\cap\overline{Q_3}}$ and $\mu_{\overline{Q_1}\cap\overline{Q_2}\cap\overline{Q_3}}$.*
The output possibility distribution is:

$$\begin{aligned}\pi^*_{b(x),3} &= \min(\pi^{*1}_{b(x)}, \pi^{*2}_{b(x)}, \pi^{*3}_{b(x)})\\ &= \min(\alpha_1,\alpha_2,\alpha_3)\mu_{Q_1\cap Q_2\cap Q_3} + \min(\beta_1,\alpha_2,\alpha_3)\mu_{\overline{Q_1}\cap Q_2\cap Q_3}\\ &\quad + \min(\alpha_1,\beta_2,\alpha_3)\mu_{Q_1\cap\overline{Q_2}\cap Q_3} + \min(\beta_1,\beta_2,\alpha_3)\mu_{\overline{Q_1}\cap\overline{Q_2}\cap Q_3}\\ &\quad + \min(\alpha_1,\alpha_2,\beta_3)\mu_{Q_1\cap Q_2\cap\overline{Q_3}} + \min(\beta_1,\alpha_2,\beta_3)\mu_{\overline{Q_1}\cap Q_2\cap\overline{Q_3}}\\ &\quad + \min(\alpha_1,\beta_2,\beta_3)\mu_{Q_1\cap\overline{Q_2}\cap\overline{Q_3}} + \min(\beta_1,\beta_2,\beta_3)\mu_{\overline{Q_1}\cap\overline{Q_2}\cap\overline{Q_3}}.\end{aligned}$$

3.2.2 Solutions for $\pi^*_{b(x),i}(u) = 1$

In the following, we study how to get $\pi^*_{b(x),i}(u) = 1$ for a value $u \in D_b$. Combining (3.5) and (3.6), we have $\pi^*_{b(x),i}(u) = N\Box^{\min}_{\max} I_i$, where N is a row of $\mathcal{M}_i$. Let us note:

$$N = \begin{bmatrix} t_1 & t_2 & \cdots & t_{2i-1} & t_{2i}\end{bmatrix} \text{ where } t_j \in \{1, r_*, s_*\}.$$

Then we have $\pi^*_{b(x),i}(u) = \min_{1\leq j\leq 2i}\max(t_j,\theta_j)$. So $\pi^*_{b(x),i}(u) = 1$ is equivalent to $\forall j \in E, \theta_j = 1$, where $E = \{j \in \{1,2,\ldots,2i\} \mid t_j < 1\}$. Thus, the normalization of the possibility distribution of b can be established by the resolution of an equation system with ω min-max equations, where the second member of at least one equation has to be equal to 1. Therefore, it is interesting to study if there are extremal solutions I_i to get $\pi^*_{b(x),i}(u) = 1$, as suggested in [55].
In what follows, we will look for an input vector column $X = [x_j]_{1\leq j\leq 2i}$, with the normalization hypothesis of its components: $\forall k \in \{1,2,\ldots,i\}$: $\max(x_{2k-1}, x_{2k}) = 1$, that satisfies the following equation:

$$N\Box^{\min}_{\max} X = 1. \tag{3.8}$$

We introduce the following order relation for solutions $X = [x_j]_{1\leq j\leq 2i}$ and $X' = [x'_j]_{1\leq j\leq 2i}$:

$$X \leq X', \text{if only if } \forall j \in \{1,2,\ldots,2i\} : x_j \leq x'_j.$$

This allows us to look for a unique minimal solution that we note S_{min} or a maximal one $S_{\max}$, if they exist.

Obviously, $S_{max} = [\theta_j^*]$ where $\theta_j^* = 1$ for each $j = 1, 2, \ldots, 2i$. For this maximal solution S_{max}, we notice that $\forall k \in \{1, 2, \ldots, i\}, (\lambda_k, \rho_k) = (1, 1)$: for each rule, the premise is considered as unknown [50].
Let us look for a unique minimal solution:
• If we have:

$$\exists k \in \{1, 2, \ldots, i\} \text{ such that } (t_{2k-1}, t_{2k}) = (1, 1), \tag{3.9}$$

then equation (3.8) does not admit a minimal solution.
• If we suppose that:

$$\forall k \in \{1, 2, \ldots, i\} \text{ we have } (t_{2k-1}, t_{2k}) \neq (1, 1), \tag{3.10}$$

then there is a unique minimal solution $S_{\min} = [\theta_j^*]_{1 \leq j \leq 2i}$ where:

$$(\theta_{2k-1}^*, \theta_{2k}^*) = \begin{cases} (1,0) & \text{if } t_{2k} = 1 \\ (0,1) & \text{if } t_{2k-1} = 1 \end{cases}.$$

So, with the hypothesis (3.10), we give a sufficient condition for the normalization of $\pi_{b(x),i}^*$.

3.2.3 Rebuild the equation system with a deleted rule

Let us delete a rule R^z from $R^1, R^2, \ldots, R^i$ and study the corresponding equation system. We denote by $B_i^{\not z}$ the matrix associated to this system of $i-1$ rules by the construction of subsection 3.1.4. We assert that the matrix $B_i^{\not z}$, which is of size $(2^{i-1}, i-1)$, can be obtained from the matrix B_i by the following practical rule:

1. We delete the z-th column of B_i. Then, we obtain a matrix of size $(2^i, i-1)$, where each row is repeated once and only once.
2. In the resulting matrix, we delete the rows $L_{k'}$ within all the pairs of rows $(L_k, L_{k'})$ where $L_k = L_{k'}$ and $k < k'$.

After these two operations, we get $B_i^{\not z}$.

To establish the pratical rule, let us take:

- the matrix B_i associated to the i rules $R^1, R^2, \ldots, R^i$ and $z \in \{1, 2, \ldots, i\}$. B_i is of the size $(2^i, i)$,
- the matrix $B_i^{\not z}$ associated to the $i-1$ rules $R^1, R^2, \ldots,$ where we deleted the rule R^z. $B_i^{\not z}$ is of size $(2^{i-1}, i-1)$ and
- the matrix B_i' of size $(2^i, i-1)$, obtained from B_i by deleting the column of index z.

We will first show the following lemma:

Lemma 1 *Each row of the matrix B_i' is repeated once and only once.*

Let us keep in B_i' a single copy of each row respecting the order of appearance of the rows (of course, we will keep the first row because it is at the top). We thus form a new matrix of size $(2^{i-1}, i-1)$ denoted B_{i-1}^*. By recurrence on i, we will prove:

Theorem 3.3 *The matrix $B_i^{\not z}$ is equal to B_{i-1}^*:*

$$B_i^{\not z} = B_{i-1}^*.$$

Such a result gives us a practical rule for obtaining the matrix $B_i^{\not z}$ and thus the matrix $M_i^{\not z}$ of the corresponding equation system.
Let us first prove Lemma 1 by recurrence on $i = 2, 3, \ldots$:

Proof 3.3 *Let us check the hypothesis for $i = 2$. If we take $z = 1$, we have:*

$$B_2 = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \alpha_2 \\ \alpha_1 & \beta_2 \\ \beta_1 & \beta_2 \end{bmatrix} \curvearrowright \begin{bmatrix} \alpha_2 \\ \alpha_2 \\ \beta_2 \\ \beta_2 \end{bmatrix}.$$

The rows L_1 and L_3 are identical to L_2 and L_4 respectively.
If we take $z = 2$, we have:

$$B_2 = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \alpha_2 \\ \alpha_1 & \beta_2 \\ \beta_1 & \beta_2 \end{bmatrix} \curvearrowright \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_1 \\ \beta_1 \end{bmatrix}.$$

The rows L_1 and L_2 are repeated in L_3 and L_4 respectively.
Let us suppose that the recurrence assumption is true up to $i-1$: in the matrix B_{i-1} (which has 2^{i-1} rows), if we delete a column of index $1 \leq z \leq i-1$, in the remaining matrix (which has 2^{i-1} rows), each row is repeated once and only once.
Let us show that it is the same for the matrix B_i. Let us delete a column of

index $1 \leq z \leq i$ *in the matrix* B_i. *We have:*

$$B_i = \left[\begin{array}{c|c} & \alpha_i \\ & \alpha_i \\ B_{i-1} & \vdots \\ & \alpha_i \\ \hline & \beta_i \\ & \beta_i \\ B_{i-1} & \vdots \\ & \beta_i \end{array}\right].$$

Let us distinguish two cases:

1. *if* $z = i$,
2. *if* $1 \leq z \leq i-1$.

1) If $z = i$, *by deleting the last column in* B_i, *the remaining matrix is:*

$$\left[\begin{array}{c} B_{i-1} \\ \hline B_{i-1} \end{array}\right].$$

We remind that in the matrix B_{i-1} *there is no row repetition. In the remaining matrix* $\left[\begin{array}{c} B_{i-1} \\ \hline B_{i-1} \end{array}\right]$, *each row is repeated one and only one time:*

$$L_1 = L_{2^{i-1}+1}\,,\, L_2 = L_{2^{i-1}+2}, \ldots, L_{2^{i-1}} = L_{2^i}$$

Therefore, we have proved the property for B_i *when we delete the column of index* $z = i$.
2) If $1 \leq z \leq i-1$, *we will consider two remaining matrices:*

- $B'_i = \left[\begin{array}{c|c|c} & & \alpha_i \\ & & \alpha_i \\ B_{i-1} & & \vdots \\ & & \alpha_i \\ \hline & & \beta_i \\ & & \beta_i \\ B_{i-1} & & \vdots \\ & & \beta_i \end{array}\right]$ *: the matrix* B_i *where the column with index* z *has been removed.*

- $B'_{i-1} = \begin{bmatrix} & | & \\ & | & \\ & | & \end{bmatrix}$: *the matrix B_{i-1} where the column with index z has been removed.*

With these notations, we get:

$$B'_i = \left[\begin{array}{c|c} & \alpha_i \\ & \alpha_i \\ B'_{i-1} & \vdots \\ & \alpha_i \\ \hline & \beta_i \\ & \beta_i \\ B'_{i-1} & \vdots \\ & \beta_i \end{array}\right].$$

Let us show that each row of B'_i is repeated once and only once.
We denote by $L_1, L_2, \ldots, L_{2^i}$ and $N_1, N_2, \ldots, N_{2^{i-1}}$ the rows of B'_i and B'_{i-1} respectively. Let k, k' be two distinct indices in $\{1, 2, \ldots, 2^i\}$:

1. *If in B'_i the row L_k coincides with the row $L_{k'}$, because of the last column of the matrix B'_i, we necessarily have:*

$$k \text{ and } k' \text{ in } \{1, 2, \ldots, 2^{i-1}\} \text{ or } k \text{ and } k' \text{ in } \{2^{i-1}+1, 2^{i-1}+2, \ldots, 2^i\}.$$

2. *If in B'_i, we have $L_k = L_{k'}$ and k, k' are in $\{1, 2, \ldots, 2^{i-1}\}$ then:*

$$L_k = (N_k, \alpha_i) \text{ and } L_{k'} = (N_{k'}, \alpha_i) \text{ with } N_k = N_{k'}.$$

   The rows $N_k, N_{k'}$ of B'_{i-1} coincide.

3. *If in B'_i, we have $L_k = L_{k'}$ and k and k' in $\{2^{i-1}+1, 2^{i-1}+2, \ldots, 2^i\}$ then:*

$$L_k = (N_{k-2^{i-1}}, \beta_i) \text{ and } L_{k'} = (N_{k'-2^{i-1}}, \beta_i) \text{ with } N_{k-2^{i-1}} = N_{k'-2^{i-1}}.$$

   The rows $N_{k-2^{i-1}}, N_{k'-2^{i-1}}$ of B'_{i-1} coincide.

But the recurrence hypothesis states that in the remaining matrix B'_{i-1} obtained by deleting the column of index z in the matrix B_{i-1}, each row repeats once and only once. We deduce by the previous considerations that in the remaining matrix B'_i obtained by deleting the column of index z in the matrix B_i, each row is repeated once and only once.

Let us now proceed to the proof of Theorem 3.3 by recurrence on $i = 2, 3, \ldots$:

Proof 3.4 *Let us check the hypothesis for $i = 2$:*

- *If we take $z = 1$, we have:*

$$B_2 = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \alpha_2 \\ \alpha_1 & \beta_2 \\ \beta_1 & \beta_2 \end{bmatrix} \curvearrowright B_2' = \begin{bmatrix} \alpha_2 \\ \alpha_2 \\ \beta_2 \\ \beta_2 \end{bmatrix} \curvearrowright B^{\not 1} = \begin{bmatrix} \alpha_2 \\ \beta_2 \end{bmatrix}, \textit{which is associated to } R^2$$

- *Otherwise, if we take $z = 2$, we have:*

$$B_2 = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \alpha_2 \\ \alpha_1 & \beta_2 \\ \beta_1 & \beta_2 \end{bmatrix} \curvearrowright B_2' = \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_1 \\ \beta_1 \end{bmatrix} \curvearrowright B^{\not 2} = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix}, \textit{which is associated to } R^1$$

Suppose that the recurrence hypothesis is true up to $i - 1$. We have:

$$B_i = \left[\begin{array}{c|c} & \alpha_i \\ & \alpha_i \\ B_{i-1} & \vdots \\ & \alpha_i \\ \hline & \beta_i \\ & \beta_i \\ B_{i-1} & \vdots \\ & \beta_i \end{array}\right]$$

We distinguish two cases:

1. *if $z = i$, by deleting the last rule R^i, by definition of the construction of the matrices B_* (subsection 3.1.4), the matrix associated to the rules R^1, R^2, R^{i-1} is B_{i-1}, so $B_i^{\not i} = B_{i-1}$. But, by deleting the column $z = i$ in the matrix B_i the remaining matrix B_i' is:*

   $$B_i' = \left[\begin{array}{c} B_{i-1} \\ \hline B_{i-1} \end{array}\right],$$

   where each row is repeated once and only once (in B_{i-1}, there is no repetition).

   If in B_i', we keep only one copy of each row (respecting the order of appearance of the rows), the obtained matrix is exactly $B_{i-1} = B_i^{\not i}$: we have proved the property for B_i and $z = i$.

2. *If $1 \leq z \leq i-1$, we will consider the following two ordered sets of rules:*

   - *The ordered set of $i-1$ rules $R^1, R^2, \ldots, R^i$ without the rule R^z, whose associated matrix by the construction in subsection 3.1.4 is denoted $B_i^{\not{z}}$.*
   - *The ordered set of $i-2$ rules $R^1, R^2, \ldots, R^{i-1}$ without the rule R^z, whose associated matrix is $B_{i-1}^{\not{z}}$.*

   In the remaining matrix $B'_{i-1} = \left[\begin{array}{c|c|c} \; & \; & \; \end{array}\right]$ obtained from B_{i-1} by deleting the column of index z, let us keep only one copy of each row (respecting the order of appearance of the rows). By the recurrence assumption, the obtained matrix is the matrix $B_{i-1}^{\not{z}}$ associated with the set of $i-2$ rules $R^1, R^2, \ldots, R^{i-1}$ (without the rule R^z).

   But, the set of $i-1$ rules $R^1, R^2, \ldots, R^i$ (without the rule R^z) is obtained from the set of $i-2$ rules $R^1, R^2, \ldots, R^{i-1}$ (without the rule R^z) by adding the rule R^i. By the construction in subsection 3.1.4, the matrix $B_i^{\not{z}}$ is:

$$
B_i^{\not{z}} = \left[\begin{array}{c|c}
 & \alpha_i \\
B_{i-1}^{\not{z}} & \alpha_i \\
 & \vdots \\
 & \alpha_i \\
\hline
 & \beta_i \\
B_{i-1}^{\not{z}} & \beta_i \\
 & \vdots \\
 & \beta_i
\end{array}\right].
$$

   Furthermore, in the remaining matrix:

$$
B'_i = \left[\begin{array}{c|c|c}
 & & \alpha_i \\
B_{i-1} & & \alpha_i \\
 & & \vdots \\
 & & \alpha_i \\
\hline
 & & \beta_i \\
B_{i-1} & & \beta_i \\
 & & \vdots \\
 & & \beta_i
\end{array}\right],
$$

which is obtained from B_i by deleting the column of index z, each row is repeated once and only once and, because of the last column, a row and its copy are both either in the top block of size $(2^{i-1}, i-1)$, or in the block below. It follows that if we keep in B'_i a single copy of each row (respecting the order of appearance of the rows), the remaining matrix is:

$$B^*_{i-1} = \left[\begin{array}{c|c} & \alpha_i \\ & \alpha_i \\ B^{\not z}_{i-1} & \vdots \\ & \alpha_i \\ \hline & \beta_i \\ & \beta_i \\ B^{\not z}_{i-1} & \vdots \\ & \beta_i \end{array}\right] = B^{\not z}_i.$$

Then, one can deduce the partition $(E^{\not z}_k)_{1\leq k\leq 2^{i-1}}$ of D_b from $B^{\not z}_i$ by using the relation (3.3). Its sets can also be determined directly from the partition $(E^{(i)}_k)_{1\leq k\leq 2^i}$. In fact, for any $k \in \{1, 2, \ldots, 2^{i-1}\}$, we can find two indices $k', k'' \in \{1, 2, \ldots, 2^i\}$ such that:

$$E^{\not z}_k = E^{(i)}_{k'} \cup E^{(i)}_{k''} \tag{3.11}$$

where, with respect to the notations of (3.3), $E^{(i)}_{k'}$ and $E^{(i)}_{k''}$ differ only on the component T_z, e.g. $T_z = Q_z$ for $E^{(i)}_{k'}$, and $T_z = \overline{Q}_z$ for $E^{(i)}_{k''}$. If $E^{\not z}_k \neq \emptyset$, the decomposition (3.11) is unique, where $E^{(i)}_{k'}$ and $E^{(i)}_{k''}$ still satisfy the above assumption. These two sets can be easily determined algorithmically.
Finally, we deduce from (3.11), that $\omega^{\not z} \leq \omega$, where $\omega^{\not z} = \mathrm{card}(J^{\not z})$ and $J^{\not z} = \Big\{k \in \{1, 2, \ldots, 2^{i-1}\} \;\Big|\; E^{\not z}_k \neq \emptyset\Big\}$.
The partition $(E^{\not z}_k)_{k\in J^{\not z}}$ and the matrix $\mathcal{B}^{\not z}_i$ are particularly interesting for performing a sensitivity analysis.

In the following, we illustrate the construction of $B^{\not z}_i$.

Example 3.3 *Let us take $i = 3$ and $z = 2$ and obtain $B_3^{\not z}$:*

$$
\begin{bmatrix}
\alpha_1 & \cancel{\alpha_2} & \alpha_3 \\
\beta_1 & \cancel{\alpha_2} & \alpha_3 \\
\alpha_1 & \cancel{\beta_2} & \alpha_3 \\
\beta_1 & \cancel{\beta_2} & \alpha_3 \\
\alpha_1 & \cancel{\alpha_2} & \beta_3 \\
\beta_1 & \cancel{\alpha_2} & \beta_3 \\
\alpha_1 & \cancel{\beta_2} & \beta_3 \\
\beta_1 & \cancel{\beta_2} & \beta_3
\end{bmatrix}
\curvearrowright
\begin{bmatrix}
\alpha_1 & \alpha_3 \\
\beta_1 & \alpha_3 \\
\cancel{\alpha_1} & \cancel{\alpha_3} \\
\cancel{\beta_1} & \cancel{\alpha_3} \\
\alpha_1 & \beta_3 \\
\beta_1 & \beta_3 \\
\cancel{\alpha_1} & \cancel{\beta_3} \\
\cancel{\beta_1} & \cancel{\beta_3}
\end{bmatrix}
\curvearrowright
\begin{bmatrix}
\alpha_1 & \alpha_3 \\
\beta_1 & \alpha_3 \\
\alpha_1 & \beta_3 \\
\beta_1 & \beta_3
\end{bmatrix}
= B_3^{\not z}.
$$

3.3 Possibility and necessity measures of any subset of the output attribute domain

In this section, we study the case of a rule-based system of n if-then possibilistic rules with its associated equation system, and denote by $\pi^*_{b(x)}$ the output possibility distribution of b.
Remind that the possibility measure Π^* and the necessity measure N^* associated to $\pi^*_{b(x)}$ are defined by [43]:

$$
\begin{aligned}
&\Pi^* : \mathcal{P}(D_b) \to [0,1] : Q \mapsto \Pi^*(Q) = \max_{u \in Q} \pi^*_{b(x)}(u), \\
&N^* : \mathcal{P}(D_b) \to [0,1] : Q \mapsto N^*(Q) = 1 - \Pi^*(\overline{Q}).
\end{aligned}
$$

For a proposition p: "$b(x) \in Q$", the possibility degree $\pi(p)$ and its degree of necessity $n(p)$ are given by [43]:

$$\pi(p) = \Pi^*(Q) \text{ and } n(p) = 1 - \pi(\neg p) = N^*(Q).$$

In what follows, we deduce explicit formulas for $\Pi^*(Q)$ and $N^*(Q)$ from (3.6), where we use a function ε, which checks if a set is not empty:

$$\varepsilon(T) = \begin{cases} 1 & \text{si } T \neq \emptyset \\ 0 & \text{if } T = \emptyset \end{cases}.$$

3.3.1 Possibility measure

For any subset $Q \subseteq D_b$, we have $Q = \bigcup_{1 \leq i \leq \omega} E^{(n)}_{k_i} \cap Q$, and we know that $\forall u \in E^{(n)}_{k_i}$ we have $\pi^*_{b(x)}(u) = o^{(n)}_{k_i}$ (subsection 3.2.1). So we get the possibility measure of Q by:

$$\Pi^*(Q) = \max_{u \in Q} \pi^*_{b(x)}(u) = \max_{1 \leq i \leq \omega \text{ s.t } E^{(n)}_{k_i} \cap Q \neq \emptyset} o^{(n)}_{k_i}.$$

Therefore we can restate this result as:

$$\begin{aligned}\Pi^*(Q) &= \max_{1 \leq i \leq \omega} \varepsilon(E^{(n)}_{k_i} \cap Q) \cdot o^{(n)}_{k_i} \\ &= \max_{1 \leq i \leq \omega} \min(\varepsilon(E^{(n)}_{k_i} \cap Q), o^{(n)}_{k_i}).\end{aligned} \tag{3.12}$$

Let ∇_Q be the matrix of size $(1, \omega)$ defined by:

$$\nabla_Q = \begin{bmatrix} \varepsilon(E^{(n)}_{k_1} \cap Q) & \varepsilon(E^{(n)}_{k_2} \cap Q) & \cdots & \varepsilon(E^{(n)}_{k_\omega} \cap Q) \end{bmatrix}.$$

Then, equality (3.12) is exactly:

$$\Pi^*(Q) = \nabla_Q \Box^{\max}_{\min} \mathcal{O}_n. \tag{3.13}$$

3.3.2 Necessity measure

Using (3.13), we have for $\overline{Q}$:

$$\Pi^*(\overline{Q}) = \nabla_{\overline{Q}} \Box^{\max}_{\min} \mathcal{O}_n.$$

The necessity measure is then:

$$N^*(Q) = 1 - \Pi^*(\overline{Q}) = (\Pi^*(\overline{Q}))^\circ.$$

By the correspondences between $\Box^{\min}_{\max}$ and $\Box^{\max}_{\min}$ we obtain:

$$N^*(Q) = (\nabla_{\overline{Q}} \Box^{\max}_{\min} \mathcal{O}_n)^\circ = \nabla_{\overline{Q}}{}^\circ \Box^{\min}_{\max} \mathcal{O}_n{}^\circ. \tag{3.14}$$

3.4 Cascade

In this section, we use two sets of if-then possibilistic rules: the n rules $R^1, R^2, \ldots, R^n$ and the m rules $R'^1, R'^2, \ldots, R'^m$. We form $\mathcal{O}_n = \mathcal{M}_n \Box^{\min}_{\max} I_n$ for the first set of rules and $\mathcal{O}'_m = \mathcal{M}'_m \Box^{\min}_{\max} I'_m$ for the second one, where we consider that their associated partition of non-empty sets $(E^{(n)}_k)_{k \in J}$ and $(E'^{(m)}_k)_{k \in J'}$ have respective size ω and ω'. In what follows, we establish an input-output relation between the two equation systems and associate to such cascade construction a min-max neural network.

3.4.1 I'_m and $\mathcal{O}'_m$ results

Each premise p'_j of a rule R'_j is a proposition of the form "$b(x) \in Q'_j$". Therefore, we get λ'_j and ρ'_j by the calculation of the possibility measures of Q'_j and $\overline{Q'_j}$:

$$\lambda'_j = \Pi^*(Q'_j) \text{ and } \rho'_j = \Pi^*(\overline{Q'_j}).$$

By the equality (3.13), we define the input vector I'_m of size $(2m, 1)$, as a max-min product between a matrix $\nabla = [\kappa_{t,k}]_{1\leq t\leq 2m, 1\leq k\leq \omega}$, and $\mathcal{O}_n$ of size $(\omega, 1)$:

$$I'_m = \nabla \Box^{max}_{min} \mathcal{O}_n. \tag{3.15}$$

The coefficients of ∇ are 1 or 0, and are obtained as follows:

$$\kappa_{t,k} = \begin{cases} \varepsilon(E_k^{(n)} \cap Q'_j) \text{ if } t = 2j-1 \text{ and } 1 \leq j \leq m \\ \varepsilon(E_k^{(n)} \cap \overline{Q'_j}) \text{ if } t = 2j \text{ and } 1 \leq j \leq m \end{cases}.$$

We note $S_1, S_2, \ldots, S_{2m}$ the rows of ∇. Each row of ∇ is in fact $\nabla_{Q'}$ associated to some set $Q' \subseteq D_b$ as in (3.13):

$$\nabla = \begin{bmatrix} \nabla_{Q'_1} \\ \nabla_{\overline{Q'_1}} \\ \vdots \\ \nabla_{Q'_m} \\ \nabla_{\overline{Q'_m}} \end{bmatrix}.$$

Thus, it establishes an input-output relation between the two equation systems. This yields the output vector $\mathcal{O}'_m$ of the second system from the first system, ∇ and $\mathcal{M}'_m$:

$$\begin{aligned} \mathcal{O}'_m &= \mathcal{M}'_m \Box^{\min}_{\max} I'_m \\ &= \mathcal{M}'_m \Box^{\min}_{\max} (\nabla \Box^{\max}_{\min} \mathcal{O}_n) \\ &= \mathcal{M}'_m \Box^{\min}_{\max} (\nabla \Box^{\max}_{\min} (\mathcal{M}_n \Box^{\min}_{\max} I_n)). \end{aligned}$$

3.4.2 Representation by a min-max neural network

In [50], the authors suggested that the system that can be built from a cascade would have a structural resemblance with a min-max neural network. We show that there is such a neural network, which gives an explicit representation of the cascade construction.

With the help of the matrices $I_n{}^\circ$, $\mathcal{M}_n{}^\circ$ and $\mathcal{M}'_m{}^\circ$, we can express the equations involved in the cascade using only the operator $(A \Box^{\max}_{\min} B)^\circ$:

$$\mathcal{O}_n = (\mathcal{M}_n{}^\circ \Box^{\max}_{\min} I_n{}^\circ)^\circ,$$

$$I'_m{}^\circ = (\nabla \square_{\min}^{\max} \mathcal{O}_n)^\circ$$

and

$$\mathcal{O}'_m = (\mathcal{M}'_m{}^\circ \square_{\min}^{\max} I'_m{}^\circ)^\circ.$$

We define the four-layer min-max neural network as follows:

- the layer 1 has $2n$ input neurons: $i_1, i_2, \ldots, i_{2n}$ with $z_{i_1}, z_{i_2}, \ldots, z_{i_{2n}}$ being their respective output values,
- the layer 2 has ω hidden neurons: $h_1, h_2, \ldots, h_\omega$ where $I_{h_1}, I_{h_2}, \ldots, I_{h_\omega}$ are their respective input values and $z_{h_1}, z_{h_2}, \ldots, z_{h_\omega}$ their respective output values,
- the layer 3 has $2m$ hidden neurons: $h'_1, h'_2, \ldots, h'_{2m}$ where $I_{h'_1}, I_{h'_2}, \ldots, I_{h'_{2m}}$ are their respective input values and $z_{h'_1}, z_{h'_2}, \ldots, z_{h'_{2m}}$ their respective output values,
- the layer 4 has ω' output neurons: $o'_1, o'_2, \ldots, o'_{\omega'}$ where $I_{o'_1}, I_{o'_2}, \ldots, I_{o'_{\omega'}}$ are their respective input values and $z_{o'_1}, z_{o'_2}, \ldots, z_{o'_{\omega'}}$ their respective output values.

In this neural network, for each neuron, we obtain its input value with the operator $(A \square_{\min}^{\max} B)^\circ$. Its output value is given by the activation function which is $f(x) = x$.

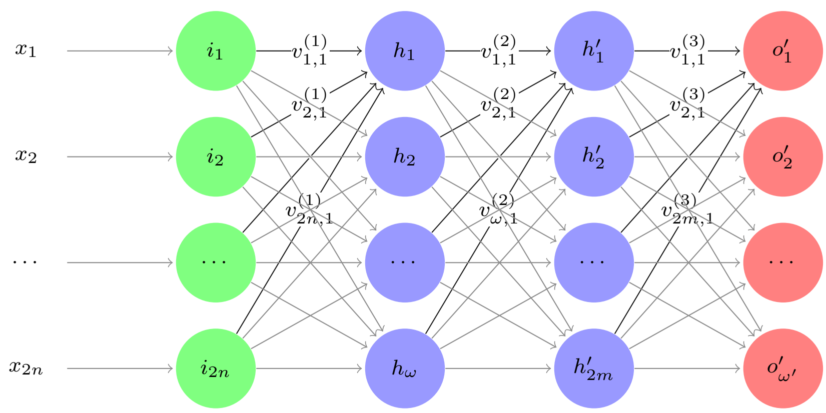


Figure 3.1: Min-max neural network architecture

We explicit its architecture (Figure 3.1) and define the following edges by:

- $x_j = 1 - \theta_j$, a coefficient of $I_n{}^\circ$ with $1 \leq j \leq 2n$,
- $v^{(1)}_{i,j} = 1 - t_{j,i}$, a coefficient of $\mathcal{M}_n{}^\circ$ with $1 \leq j \leq \omega$ and $1 \leq i \leq 2n$,
- $v^{(2)}_{i,j} = \kappa_{j,i}$, a coefficient of ∇ with $1 \leq j \leq 2m$ and $1 \leq i \leq \omega$,

- $v^{(3)}_{i,j} = 1 - t'_{j,i}$, a coefficient of ${\mathcal{M}'_m}^\circ$ with $1 \leq j \leq \omega'$ and $1 \leq i \leq 2m$.

The output value of an input neuron i_k is $z_{i_k} = f(x_k) = 1 - \theta_k$. For each hidden neuron h_k, its output value $z_{h_k} = f(I_{h_k}) = I_{h_k}$ is a coefficient of $\mathcal{O}_n$, as I_{h_k} is obtained using a row N_k of $\mathcal{M}_n$:

$$I_{h_k} = 1 - \max_{1 \leq j \leq 2n} \min(v^{(1)}_{j,k}, z_{i_j}) = N_k \square^{\min}_{\max} I_n.$$

Each output value of a hidden neuron h'_k is a coefficient of ${I'_m}^\circ$. We use the row S_k of ∇ to obtain $I_{h'_k}$:

$$I_{h'_k} = 1 - \max_{1 \leq j \leq \omega} \min(v^{(2)}_{j,k}, z_{h_j}) = (S_k \square^{\max}_{\min} O_n)^\circ.$$

We have $z_{h'_k} = f(I_{h'_k}) = I_{h'_k}$. Finally, $I_{o'_k}$ corresponding to the output neuron o'_k is obtained using the row N'_k of $\mathcal{M}'_m$:

$$I_{o'_k} = 1 - \max_{1 \leq j \leq 2m} \min(v^{(3)}_{j,k}, z_{h'_j}) = N'_k \square^{\min}_{\max} I'_m.$$

We get the output value of o'_k with $z_{o'_k} = f(I_{o'_k}) = I_{o'_k}$. So $z_{o'_1}, z_{o'_2}, \ldots, z_{o'_{\omega'}}$ are the coefficients of $\mathcal{O}'_m$.
As characteristics, we notice that each edge $v^{(2)}_{i,j}$ has a value equal to 0 or 1 with respect to the relation (3.15), while the values of the others are in $[0, 1]$. Furthermore, it has some resemblance with an hybrid fuzzy neural network [28], where the t-norm min and its associated t-conorm max are used to get the input value of a neuron. By using more layers, we can extend this min-max neural network to take into account the λ, ρ calculations when the premises are compounded.

3.4.3 Example

To illustrate the cascade, we use the example of [50], previously introduced in the french version of [55]. It is a possibilistic rule-based system which suggests to people professions with associated salaries, based on their tastes and interests using two sets of if-then possibilistic rules.
Firstly, the inference of three possibilistic rules determine which professions can be suggested to a person, according to her characteristics:

- R^1: if a person likes meeting people, then recommended professions are professor or business man or lawyer or doctor,

- R^2: if a person is fond of creation/invention, then recommended professions are engineer or architect,

- R^3: if a person looks for job security and is fond of intellectual speculation, then recommended professions are professor or researcher.

The parameters of the rules are $s_1 = 1$, $r_1 = 0.3$, $s_2 = 0.2$, $r_2 = 0.4$, $s_3 = 1$, $r_3 = 0.3$. These rules are constructed from four input attributes "likes meeting people", "fond of creation/invention", "looks for job security", "fond of intellectual speculation" and one output attribute "profession". The attribute domains are: $D_{\text{likes meeting people}} = D_{\text{fond of creation/invention}} = D_{\text{looks for job security}} = D_{\text{fond of intellectual speculation}} = \{\text{yes, no}\}$ and $D_{\text{profession}} = \{$professor, business man, lawyer, doctor, engineer, architect, researcher, others$\}$.
For this set of rules, an equation system is formed, where $\mathcal{O}_n$ has five coefficients. In fact, the eight possible professions are in five non-empty disjoint sets which form a partition: $E^{(3)}_{k_1} = \{\text{researcher}\}$, $E^{(3)}_{k_2} = \{\text{professor}\}$, $E^{(3)}_{k_3} = \{\text{engineer, architect}\}$, $E^{(3)}_{k_4} = \{\text{business man, lawyer, doctor}\}$ and $E^{(3)}_{k_5} = \{\text{others}\}$. With the possibility distributions of the input attributes of [55], we get $\lambda_1 = 1, \rho_1 = 0.5, \lambda_2 = 0.2, \rho_2 = 1, \lambda_3 = 1, \rho_3 = 0.6$. We form the equation system $\mathcal{O}_n = \mathcal{M}_n \Box^{\min}_{\max} I_n$ and perform the matricial products:

$$\begin{bmatrix} \Pi(E^{(3)}_{k_1}) \\ \Pi(E^{(3)}_{k_2}) \\ \Pi(E^{(3)}_{k_3}) \\ \Pi(E^{(3)}_{k_4}) \\ \Pi(E^{(3)}_{k_5}) \end{bmatrix} = \begin{bmatrix} 1 & r_1 & s_2 & 1 & s_3 & 1 \\ s_1 & 1 & 1 & r_2 & s_3 & 1 \\ 1 & r_1 & s_2 & 1 & 1 & r_3 \\ s_1 & 1 & 1 & r_2 & 1 & r_3 \\ 1 & r_1 & 1 & r_2 & 1 & r_3 \end{bmatrix} \Box^{\min}_{\max} \begin{bmatrix} \lambda_1 \\ \rho_1 \\ \lambda_2 \\ \rho_2 \\ \lambda_3 \\ \rho_3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 1 \\ 0.2 \\ 0.6 \\ 0.5 \end{bmatrix}.$$

We form $\mathcal{O}_n = \boxdot_{\min} \mathcal{B}_n$:

$$\begin{bmatrix} \Pi(E^{(3)}_{k_1}) \\ \Pi(E^{(3)}_{k_2}) \\ \Pi(E^{(3)}_{k_3}) \\ \Pi(E^{(3)}_{k_4}) \\ \Pi(E^{(3)}_{k_5}) \end{bmatrix} = \boxdot_{\min} \begin{bmatrix} \beta_1 & \alpha_2 & \alpha_3 \\ \alpha_1 & \beta_2 & \alpha_3 \\ \beta_1 & \alpha_2 & \beta_3 \\ \alpha_1 & \beta_2 & \beta_3 \\ \beta_1 & \beta_2 & \beta_3 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 1 \\ 0.2 \\ 0.6 \\ 0.5 \end{bmatrix}.$$

Then, based on this result, the system determines the salary she can expect according to her profession, using three rules:

- R'^1: if a person is a professor or a researcher, then her salary is low,
- R'^2: if a person is an engineer, a lawyer or an architect then her salary is average or high,
- R'^3: if a person is a business man or a doctor, then her salary is high.

For the attribute salary, its domain is $D_{\text{salary}} = \{\text{low, average, low}\}$ and the sets $E'^{(3)}_{k_1} = \{\text{high}\}$, $E'^{(3)}_{k_2} = \{\text{average}\}$ and $E'^{(3)}_{k_3} = \{\text{low}\}$ form the partition of its domain. So $\mathcal{O}'_m$, $\mathcal{M}'_m$ and $\mathcal{B}'_m$ have three rows. Using the partition of the first system and the sets within the propositions p'_1, p'_2 and p'_3 of the three rules R'_1, R'_2 and R'_3 respectively, we get ∇:

$$\nabla = \begin{bmatrix} \nabla_{Q'_1} \\ \nabla_{\overline{Q'_1}} \\ \nabla_{Q'_2} \\ \nabla_{\overline{Q'_2}} \\ \nabla_{Q'_3} \\ \nabla_{\overline{Q'_3}} \end{bmatrix}$$

$$= \begin{bmatrix} \varepsilon(E^{(3)}_{k_1} \cap Q_1) & \varepsilon(E^{(3)}_{k_2} \cap Q_1) & \varepsilon(E^{(3)}_{k_3} \cap Q_1) & \varepsilon(E^{(3)}_{k_4} \cap Q_1) & \varepsilon(E^{(3)}_{k_5} \cap Q_1) \\ \varepsilon(E^{(3)}_{k_1} \cap \overline{Q_1}) & \varepsilon(E^{(3)}_{k_2} \cap \overline{Q_1}) & \varepsilon(E^{(3)}_{k_3} \cap \overline{Q_1}) & \varepsilon(E^{(3)}_{k_4} \cap \overline{Q_1}) & \varepsilon(E^{(3)}_{k_5} \cap \overline{Q_1}) \\ \varepsilon(E^{(3)}_{k_1} \cap Q_2) & \varepsilon(E^{(3)}_{k_2} \cap Q_2) & \varepsilon(E^{(3)}_{k_3} \cap Q_2) & \varepsilon(E^{(3)}_{k_4} \cap Q_2) & \varepsilon(E^{(3)}_{k_5} \cap Q_2) \\ \varepsilon(E^{(3)}_{k_1} \cap \overline{Q_2}) & \varepsilon(E^{(3)}_{k_2} \cap \overline{Q_2}) & \varepsilon(E^{(3)}_{k_3} \cap \overline{Q_2}) & \varepsilon(E^{(3)}_{k_4} \cap \overline{Q_2}) & \varepsilon(E^{(3)}_{k_5} \cap \overline{Q_2}) \\ \varepsilon(E^{(3)}_{k_1} \cap Q_3) & \varepsilon(E^{(3)}_{k_2} \cap Q_3) & \varepsilon(E^{(3)}_{k_3} \cap Q_3) & \varepsilon(E^{(3)}_{k_4} \cap Q_3) & \varepsilon(E^{(3)}_{k_5} \cap Q_3) \\ \varepsilon(E^{(3)}_{k_1} \cap \overline{Q_3}) & \varepsilon(E^{(3)}_{k_2} \cap \overline{Q_3}) & \varepsilon(E^{(3)}_{k_3} \cap \overline{Q_3}) & \varepsilon(E^{(3)}_{k_4} \cap \overline{Q_3}) & \varepsilon(E^{(3)}_{k_5} \cap \overline{Q_3}) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Then, we form $I'_m = \nabla \Box^{max}_{min} \mathcal{O}_n$:

$$I'_m = \begin{bmatrix} \lambda'_1 \\ \rho'_1 \\ \lambda'_2 \\ \rho'_2 \\ \lambda'_3 \\ \rho'_3 \end{bmatrix} = \begin{bmatrix} \nabla_{Q'_1} \\ \nabla_{\overline{Q'_1}} \\ \nabla_{Q'_2} \\ \nabla_{\overline{Q'_2}} \\ \nabla_{Q'_3} \\ \nabla_{\overline{Q'_3}} \end{bmatrix} \Box^{max}_{min} \begin{bmatrix} \Pi(E^{(3)}_{k_1}) \\ \Pi(E^{(3)}_{k_2}) \\ \Pi(E^{(3)}_{k_3}) \\ \Pi(E^{(3)}_{k_4}) \\ \Pi(E^{(3)}_{k_5}) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.6 \\ 0.6 \\ 1 \\ 0.6 \\ 1 \end{bmatrix}.$$

We arbitrarily set $s'_1 = 1$, $r'_1 = 0.7$, $s'_2 = 0.8$, $r'_2 = 0.2$, $s'_3 = 0.6$ and $r'_3 = 0.4$. Thus, we now form $\mathcal{O}'_m = \mathcal{M}'_m \Box^{\min}_{\max} I'_m$ and perform the matricial products:

$$\begin{bmatrix} \Pi(E'^{(3)}_{k_1}) \\ \Pi(E'^{(3)}_{k_2}) \\ \Pi(E'^{(3)}_{k_3}) \end{bmatrix} = \begin{bmatrix} 1 & r'_1 & s'_2 & 1 & s'_3 & 1 \\ 1 & r'_1 & s'_2 & 1 & 1 & r'_3 \\ s'_1 & 1 & 1 & r'_2 & 1 & r'_3 \end{bmatrix} \Box^{\min}_{\max} \begin{bmatrix} \lambda'_1 \\ \rho'_1 \\ \lambda'_2 \\ \rho'_2 \\ \lambda'_3 \\ \rho'_3 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.7 \\ 1 \end{bmatrix}.$$

We form $\mathcal{O}'_m = \boxdot_{\min} \mathcal{B}'_m$:

$$\begin{bmatrix} \Pi(E'^{(3)}_{k_1}) \\ \Pi(E'^{(3)}_{k_2}) \\ \Pi(E'^{(3)}_{k_3}) \end{bmatrix} = \boxdot_{\min} \begin{bmatrix} \beta'_1 & \alpha'_2 & \alpha'_3 \\ \beta'_1 & \alpha'_2 & \beta'_3 \\ \alpha'_1 & \beta'_2 & \beta'_3 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.7 \\ 1 \end{bmatrix}.$$

Such cascade is represented by a min-max neural network (Figure 3.2).

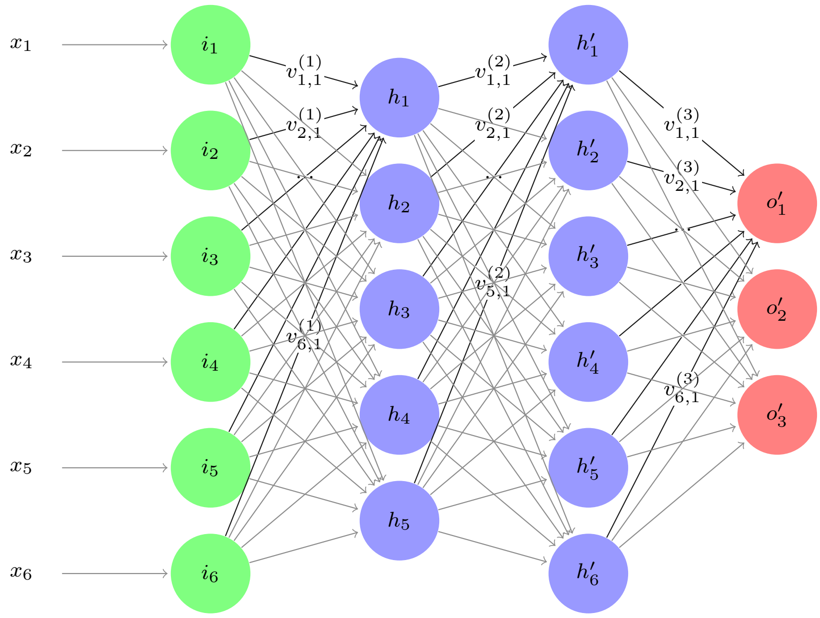


Figure 3.2: Min-max neural network architecture for the example.

The edges are computed with the help of the following matrices:

$$I_n{}^\circ = \begin{bmatrix} 0 \\ 0.5 \\ 0.8 \\ 0 \\ 0 \\ 0.4 \end{bmatrix},$$

$$\mathcal{M}_n{}^\circ = \begin{bmatrix} 0 & 0.7 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0 \\ 0 & 0.7 & 0.8 & 0 & 0 & 0.7 \\ 0 & 0 & 0 & 0.6 & 0 & 0.7 \\ 0 & 0.7 & 0 & 0.6 & 0 & 0.7 \end{bmatrix} \text{ and}$$

$$\mathcal{M}'_m{}^\circ = \begin{bmatrix} 0 & 0.3 & 0.2 & 0 & 0.4 & 0 \\ 0 & 0.3 & 0.2 & 0 & 0 & 0.6 \\ 0 & 0 & 0 & 0.8 & 0 & 0.6 \end{bmatrix}.$$

The output value of the input neurons $i_1, i_2, \ldots, i_6$ are $z_{i_1} = f(x_1) = 0$, $z_{i_2} = 0.5$, $z_{i_3} = 0.8$, $z_{i_4} = 0$, $z_{i_5} = 0$ and $z_{i_6} = 0.4$. The output value of the hidden neurons h_1, h_2, h_3, h_4 and h_5 are $z_{h_1} = f(I_{h_1}) = I_{h_1}$, $z_{h_2} = I_{h_2}$, $z_{h_3} = I_{h_3}$, $z_{h_4} =$

I_{h_4} and $z_{h_5} = I_{h_5}$. We compute $I_{h_1}, I_{h_2} \dots, I_{h_5}$ as follows:

$$I_{h_1} = 1 - \max_{1\leq j\leq 6} \min(v^{(1)}_{j,1}, z_{i_j}) = 1 - \max(0.5, 0.8) = 0.2.$$
$$I_{h_2} = 1.$$
$$I_{h_3} = 0.2.$$
$$I_{h_4} = 0.6.$$
$$I_{h_5} = 0.5.$$

So, $z_{h_1}, z_{h_2}, z_{h_3}, z_{h_4}$ and z_{h_5} are the coefficients of $\mathcal{O}_n$. We compute $I_{h'_1}, I_{h'_2} \dots, I_{h'_6}$ for the hidden neurons $h'_1, h'_2, \dots, h'_6$:

$$I_{h'_1} = 1 - \max_{1\leq j\leq 5} \min(v^{(2)}_{j,1}, z_{h_j}) = 1 - \max(0.2, 1) = 0.$$
$$I_{h'_2} = 0.4.$$
$$I_{h'_3} = 0.4.$$
$$I_{h'_4} = 0.$$
$$I_{h'_5} = 0.4.$$
$$I_{h'_6} = 0.$$

The output value $z_{h'_1} = f(I_{h'_1}) = I_{h'_1} = 0, z_{h'_2} = 0.4, z_{h'_3} = 0.4, z_{h'_4} = 0, z_{h'_5} = 0.4$ and $z_{h'_6} = 0$ of the the hidden neurons $h'_1, h'_2, \dots, h'_6$ are the coefficients of ${I'_m}^\circ$. Finally, we compute $I_{o'_1}$,$I_{o'_2}$ and $I_{o'_3}$, which correspond to the output neurons o'_1, o'_2 and o'_3 respectively:

$$I_{o'_1} = 1 - \max_{1\leq j\leq 6} \min(v^{(3)}_{j,1}, z_{h'_j}) = 1 - \max(0.3, 0.2, 0, 4) = 0.6.$$
$$I_{o'_2} = 0.7.$$
$$I_{o'_3} = 1.$$

So, $z_{o'_1} = f(I_{o'_1}) = I_{o'_1} = 0.6, z_{o'_2} = 0.7$ and $z_{o'_3} = 1$ are the coefficients of $\mathcal{O}'_m$.

3.5 Conclusion

In this chapter, we gave a canonical construction for the matrices governing the min-max equation system of Farreny and Prade [55]. This equation system was proposed for developing the explanatory capacities of possibilistic rule-based system. From our generalized equation system, we obtained an explicit formula for the output possibility distribution and computed the corresponding possibility and necessity measures. We gave a necessary and sufficient condition for the output possibility distribution to be normalized and determined, when it is possible, minimal input solutions for the normalization. We defined an algorithm to rebuild the equation system when we remove a rule.

It outputs the equation system associated to the remaining subset of rules. Therefore, this algorithm enables us to obtain all the equation subsystems of an initial equation system.

We extended our equation system for the case of a cascade. We have defined an input-output relation between the equation systems associated to each set of rules: it links the output vector of the first system to the input vector of the second system. Therefore, the output vector of the second system is described by nested min-max products of the matrices of the two equation systems. Finally, we have shown that our cascade construction can be represented by an explicit min-max neural network.

PART C

Justification of the inference results of fuzzy and possibilistic rule-based systems

In this part, we present methods for extracting the content of explanations, which constitute the first step of our processing chain for the production of natural language explanations (Figure 1). We justify the inference results of two rule-based systems: a possibilistic rule-based system (Chapter 4) and a fuzzy rule-based system composed of possibility rules (Mamdani fuzzy inference system) (Chapter 5). For the case of a possibilistic rule-based system, we rely on an early work by Farreny and Prade on the explainability of such system [55], that we reminded in Chapter 2.

For both types of rule-based systems, we start by selecting the rule premises that justify an inference result. Then, we introduce premise reduction functions and apply them to the selected premises. This allows us to form two explanations: the *justification* and the *unexpectedness* of an inference result. The justification is a set of logical expressions sufficient to justify the considered inference result. The unexpectedness is a set of logical expressions, which are not involved in the determination of the considered inference result, although they may appear to be a potential incompatibility between them and the considered inference result.

In Part D, a representation of these explanations in terms of conceptual graphs will be defined.

Chapter 4

Justification of possibilistic rule-based system inference results

The work in this chapter has led to the publication of a conference paper: Baaj, I., Poli, J. P., Ouerdane, W. & Maudet, N. (2021, September). Representation of Explanations of Possibilistic Inference Decisions. In 2021 European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU).

In this chapter, we tackle the explainability of the inference results of a possibilistic rule-based system. We rely on the Farreny and Prade's approach [55], which was reminded in Chapter 2. Using their min-max equation system, the authors study 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. The authors of [55] give a sufficient condition to obtain $\pi^*_{b(x)}(u) > \tau$ for a particular pair (u,τ) of their example. For the second question, they 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, after introducing some notations and an example that will be used to illustrate all our constructions (Section 4.1), we address these two questions in the general case. For the first question, we give necessary and sufficient conditions to obtain $\pi^*_{b(x)}(u) > \tau$ and $\pi^*_{b(x)}(u) < \tau$ according to the degrees of premises (Section 4.2). For the second question, we give a necessary and sufficient condition that allows us to justify $\pi^*_{b(x)}(u) = \tau$ by degrees of

premises (Section 4.3). This allows to extract the subset of premises whose degrees are involved in the computation of $\pi^*_{b(x)}(u)$ (Section 4.4).
We then define four premise reduction functions (Section 4.5) and apply them to the obtained subset of premises related to $\pi^*_{b(x)}(u)$. This leads us to form two kinds of explanations of $\pi^*_{b(x)}(u)$ (Section 4.6):

- The *justification* of $\pi^*_{b(x)}(u)$, which 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* of $\pi^*_{b(x)}(u)$, which 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.

Finally, we illustrate our constructions by another example (Section 4.7) and conclude with some perspectives (Section 4.8).

4.1 Notations

In this section, we introduce some notations that will be useful for studying the explainability of possibilistic rule-based systems. For a set output attribute value $u \in D_b$, the computation of its possibility degree is given by:

$$\pi^*_{b(x)}(u) = \min(\gamma_1, \gamma_2, \dots, \gamma_n), \tag{4.1}$$

$$\text{where } \gamma_i = \pi^{*i}_{b(x)}(u) = \max(t_i, \theta_i) \text{ with } (t_i, \theta_i) = \begin{cases} (s_i, \lambda_i) & \text{if } \gamma_i = \alpha_i \\ (r_i, \rho_i) & \text{if } \gamma_i = \beta_i \end{cases}. \tag{4.2}$$

The relation (4.1) is a more convenient formulation of (1.2). According to (4.2), for each $i = 1, 2, \dots, n$, 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 $\gamma_1, \gamma_2, \ldots, \gamma_n$ appearing in the relation (4.1). For $i = 1, 2, \ldots, n$, we set:

$$(p_i, sem_i, d_i) = \begin{cases} (p_i, \mathsf{P}, \lambda_i) & \text{if} \quad \gamma_i = \alpha_i \\ (p_i, \mathsf{C}, 1 - \rho_i) & \text{if} \quad \gamma_i = \beta_i \end{cases}. \tag{4.3}$$

Our notations are illustrated by the following example of a possibilistic rule-based system:

Example 4.1 *Possibilistic rule-based systems have been used in medecine, for instance DIABETO [29] enables an improvement in the dietetics of diabetic patients [99]. We propose a possibilistic rule-based system for controlling the blood sugar level of a patient with type 1 diabetes (Table 4.1), according to some factors [27]:*

	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 4.1: 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}.
By the relation (4.1)*, the possibility degree of the three output attribute values low, medium and high are:*

- $\pi_{fbs(x)}(low) = \min(\gamma_1^l, \gamma_2^l, \gamma_3^l)$*, where* $\gamma_1^l = \beta_1 = \max(r_1, \rho_1)$, $\gamma_2^l = \alpha_2 = \max(s_2, \lambda_2)$ *and* $\gamma_3^l = \alpha_3 = \max(s_3, \lambda_3)$.
- $\pi_{fbs(x)}(medium) = \min(\gamma_1^m, \gamma_2^m, \gamma_3^m)$*, where* $\gamma_1^m = \beta_1$, $\gamma_2^m = \beta_2 = \max(r_2, \rho_2)$ *and* $\gamma_3^m = \alpha_3$.
- $\pi_{fbs(x)}(high) = \min(\gamma_1^h, \gamma_2^h, \gamma_3^h)$*, where* $\gamma_1^h = \alpha_1 = \max(s_1, \lambda_1)$, $\gamma_2^h = \beta_2$ *and* $\gamma_3^h = \beta_3 = \max(r_3, \rho_3)$.

Using Notation 1, the following triplets are set (relation 4.3):

- *for* $\pi_{fbs(x)}(low)$*:* $(p_1, C, 1 - \rho_1), (p_2, P, \lambda_2)$ *and* (p_3, P, λ_3).
- *for* $\pi_{fbs(x)}(medium)$*:* $(p_1, C, 1 - \rho_1), (p_2, C, 1 - \rho_2)$ *and* (p_3, P, λ_3).
- *for* $\pi_{fbs(x)}(high)$*:* $(p_1, P, \lambda_1), (p_2, C, 1 - \rho_2)$ *and* $(p_3, C, 1 - \rho_3)$.

We give an example of inference of this blood sugar control system:

Example 4.2 *In our example 4.1, as parameters of the rules, we take:*

$$s_1 = 1, \quad s_2 = 0.7, \quad s_3 = 1 \quad \textit{and} \quad r_1 = r_2 = r_3 = 0.$$

The three rules are certain [53] because we have $\pi(q_i \mid p_i) = 1$ *and* $r_i = \pi(\neg q_i \mid p_i) = 0$. *Moreover, we assume that:*

$$\pi_{act(x)}(\textit{drink-coffee}) = 1, \quad \pi_{cbs(x)}(\textit{medium}) = 1 \quad \textit{and} \quad \pi_{cbs(x)}(\textit{low}) = 0.3$$

while the other elements of the domains of the input attributes have a possibility degree equal to zero. Then, by easy computations we get:

$$\lambda_1 = 1, \rho_1 = 0.3, \lambda_2 = 0, \rho_2 = 1, \lambda_3 = 0 \textit{ and } \rho_3 = 1.$$

The obtained output possibility distribution is:

$$\pi^*_{fbs(x)}(\textit{low}) = 0.3, \quad \pi^*_{fbs(x)}(\textit{medium}) = 0.3 \quad \textit{and} \quad \pi^*_{fbs(x)}(\textit{high}) = 1.$$

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

Taking advantage of the notations (4.1) and (4.2), we note that $\pi^*_{b(x)}(u)$ ranges between $\omega = \min(t_1, t_2, \ldots, t_n)$ and 1. Following this, a necessary and sufficient condition to obtain $\pi^*_{b(x)}(u) < \tau$ with $\omega < \tau \leq 1$ according to the degrees of premises can be easily stated:

$$\exists i \in \Big\{ j \in \{1, 2, \ldots, n\} \mid t_j < \tau \Big\} \text{ such that } \theta_i < \tau.$$

$\theta_i < \tau$ is achieved by the following condition on the degrees of the propositions composing the premise $p_i = p^i_1 \wedge p^i_2 \wedge \cdots \wedge p^i_k$ of the rule R^i, depending on whether $\gamma_i = \alpha_i$ or $\gamma_i = \beta_i$:

- If $\gamma_i = \alpha_i$, the condition $\theta_i < \tau$ relates on the possibility degree of the premise p_i, which has to be strictly lower than τ:

$$\exists j \in \{1, 2, \ldots, k\} \text{ such that } \pi(p^i_j) < \tau.$$

- If $\gamma_i = \beta_i$, the condition $\theta_i < \tau$ relates on the possibility degree of the negation of the premise p_i. The negation of each proposition composing p_i must have a possibility degree strictly lower than τ:

$$\forall j \in \{1, 2, \ldots, k\}, \pi(\neg p^i_j) < \tau.$$

Similarly, a necessary and sufficient condition to obtain $\pi^*_{b(x)}(u) > \tau$ with $\omega \leq \tau < 1$ is:

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

$\theta_i > \tau$ is achieved by conditions on the propositions composing the premise $p_i = p^i_1 \wedge p^i_2 \wedge \cdots \wedge p^i_k$, according to γ_i:

- If $\gamma_i = \alpha_i$, the condition $\theta_i > \tau$ relates on the possibility degree of the premise p_i. The possibility degree of each proposition composing p_i must be strictly greater than τ:

$$\forall j \in \{1, 2, \dots, k\}, \pi(p^i_j) > \tau.$$

- If $\gamma_i = \beta_i$, the condition $\theta_i > \tau$ relates on the possibility degree of the negation of the premise p_i. There exists a proposition composing p_i, whose possibility degree of its negation is strictly greater than τ:

$$\exists j \in \{1, 2 \dots, k\} \text{ such that } \pi(\neg p^i_j) > \tau.$$

The above conditions related to the propositions composing the premises of the rules can be achieved in terms of the values of the possibility distributions of the input attributes. For instance, for a proposition p^i_j of the form "$a^i_j(x) \in P^i_j$", the condition $\pi(p^i_j) < \tau$ is equivalent to $\forall v \in P^i_j, \pi_{a^i_j(x)}(v) < \tau$.

Example 4.3 *In our example 4.1, suppose we want to have $\pi^*_{fbs(x)}(high) > 0.5$. If we assume that:*

$$s_1 = 1, \quad r_2 = 0 \quad \textit{and} \quad r_3 = 0,$$

we have $\pi^*_{fbs(x)}(high) > 0.5$ iff $\rho_2 > 0.5$ and $\rho_3 > 0.5$.

To have $\rho_2 > 0.5$, the negation of the proposition "act(x) $\in$ {long-sleep, sport, walking}" or the negation of "cbs(x) $\in$ {low, medium}" must have a possibility degree strictly greater than 0.5.
Similarly, to have $\rho_3 > 0.5$, the negation of "act(x) $\in$ {alcohol-consumption, breakfast}" or the negation of "cbs(x) $\in$ {low, medium}" must have a possibility degree strictly greater than 0.5.
We note that if the negation of "cbs(x) $\in$ {low, medium}" has a possibility degree strictly greater than 0.5, then we have $\pi_{cbs(x)}(high) > 0.5$.

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

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 $t_1, t_2, \dots t_n$ of the rules to the degrees $\theta_1, \theta_2, \dots, \theta_n$ of the premises in the relation (4.1). Intuitively, J^P (resp. J^R) collects indices where θ_i is greater (resp. lower) than t_i: in other words, $\gamma_i = \pi^{*i}_{b(x)}(u)$ which is clearly related to the rule R^i, can be explained by a degree of the premise (resp. by a parameter of the rule):

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

We have $\{1, 2, \ldots, n\} = J^P \cup J^R$ but J^P or J^R may be empty. With the convention $\min_\emptyset = 1$, we take:

$$c_\theta = \min_{i \in J^P} \theta_i \quad \text{and} \quad c_t = \min_{i \in J^R} t_i. \tag{4.4}$$

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).
With the notations (4.4) and using the properties of the min function, we establish:

Proposition 4.1

$$\tau = \min(c_\theta, c_t). \tag{4.5}$$

Proof 4.1 *We prove $\tau = \min(c_\theta, c_t)$ in two steps:*

1. $\tau \leq \min(c_\theta, c_t)$,
2. $\tau \geq \min(c_\theta, c_t)$.

For the two steps, we remind that we have

$$\tau = \min(\gamma_1, \gamma_2, \ldots, \gamma_n)$$

1) For the first step, it is equivalent to show that:

$\tau \leq c_\theta$ *and* $\tau \leq c_t$.

- $\tau \leq c_\theta$:
  If $c_\theta = 1$, *this inequality is clear because* $\tau \leq 1$.
  Otherwise, if $c_\theta < 1$ *we have:*
  - $J^P \neq \emptyset$ *as* $\min_\emptyset = 1$.
  - *There exists an index* $i_0 \in J^P$ *such that:*

    $$c_\theta = \min_{i \in J^P} \theta_i = \theta_{i_0}.$$

    By definition of the set J^P, *we deduce:*

    $$t_{i_0} \leq \theta_{i_0}.$$

    Therefore, we have:

    $$\gamma_{i_0} = \max(t_{i_0}, \theta_{i_0}) = \theta_{i_0}.$$

Finally, we obtain:

$$\tau \leq \gamma_{i_0} = \theta_{i_0} = c_\theta.$$

- $\tau \leq c_t$:
  If $c_t = 1$, *this inequality is clear because* $\tau \leq 1$.
  Otherwise, if $c_t < 1$ *we have:*

  - $J^R \neq \emptyset$ *as* $\min_\emptyset = 1$.
  - *There exists an index* $i_0 \in J^R$ *such that:*

    $$c_t = \min_{i \in J^R} t_i = t_{i_0}.$$

    By definition of the set J^R, *we deduce:*

    $$\theta_{i_0} \leq t_{i_0}.$$

    Therefore, we have:

    $$\gamma_{i_0} = \max(t_{i_0}, \theta_{i_0}) = t_{i_0}.$$

    Finally, we obtain:

    $$\tau \leq \gamma_{i_0} = t_{i_0} = c_t.$$

2) For the second step, as we have

$$\tau = \min(\gamma_1, \gamma_2, \ldots, \gamma_n)$$

there exists an index $i_0 \in \{1, 2, \ldots, n\}$ *such that:*

$$\tau = \min(\gamma_1, \gamma_2, \ldots, \gamma_n) = \gamma_{i_0}.$$

As we have $\{1, 2, \ldots, n\} = J^P \cup J^R$, *then either* $i_0 \in J^P$ *or* $i_0 \in J^R$.
In both cases $i_0 \in J^P$ *or* $i_0 \in J^R$, *we will show the intended inequality* $\tau \geq \min(c_\theta, c_t)$:

- *If* $i_0 \in J^P$, *we have* $t_{i_0} \leq \theta_{i_0}$ *and:*

  - $c_\theta = \min_{i \in J^P} \theta_i \leq \theta_{i_0}$,
  - $\gamma_{i_0} = \max(t_{i_0}, \theta_{i_0}) = \theta_{i_0}$.

  By definition of the index i_0, *we finally have:*

  $$\tau = \gamma_{i_0} = \theta_{i_0} \geq c_\theta \geq \min(c_\theta, c_t).$$

- *If* $i_0 \in J^R$*, we have* $\theta_{i_0} \leq t_{i_0}$ *and:*
  - $c_t = \min_{i \in J^R} t_i \leq t_{i_0}$,
  - $\gamma_{i_0} = \max(t_{i_0}, \theta_{i_0}) = t_{i_0}$.

By definition of the index i_0*, we finally have:*

$$\tau = \gamma_{i_0} = t_{i_0} \geq c_t \geq \min(c_\theta, c_t).$$

As suggested above, *it may happen that we cannot justify* $\pi^*_{b(x)}(u) = \tau$ *by degrees of premises.* In fact, as the degrees $\theta_1, \theta_2, \dots, \theta_n$ of the premises are computed using the possibility distributions of the input attributes, we may have $J^P = \emptyset$. In that case, $c_\theta = 1$, $J^R = \{1, 2, \dots, n\}$ and:

$$\pi^*_{b(x)}(u) = c_t = \min(t_1, t_2, \dots, t_n). \tag{4.6}$$

Clearly, it appears that $\pi^*_{b(x)}(u)$ is independent from $\theta_1, \theta_2, \dots, \theta_n$ and we cannot justify $\pi^*_{b(x)}(u) = \tau$ by degrees of premises.

Example 4.4 *With the hypotheses of example 4.2, we form the following sets using the relation (4.4) for each output attribute value:*

- *for low:* $J^P_l = \{1\}$ *and* $J^R_l = \{2, 3\}$,
- *for medium:* $J^P_m = \{1, 2\}$ *and* $J^R_m = \{3\}$,
- *for high:* $J^P_h = \{1, 2, 3\}$ *and* $J^R_h = \{1\}$.

Then, we deduce for each output attribute value:

- *for low:* $c^l_\theta = 0.3 = \pi_{fbs(x)}(low)$ *and* $c^l_t = 0.7$.
- *for medium:* $c^m_\theta = 0.3 = \pi_{fbs(x)}(medium)$ *and* $c^m_t = 1$.
- *for high:* $c^h_\theta = c^h_t = 1 = \pi_{fbs(x)}(high)$.

4.4 Extraction of premises justifying $\pi^*_{b(x)}(u) = \tau$

To explain the inference results of our possibilistic rule-based system, we introduce a threshold $\eta > 0$. Such threshold is set according to what is modeled by the rule-base and has the following purpose:

Definition 4.1 *If a possibility (resp. necessity) degree is higher than the threshold* η*, it intuitively means that the information it models is relevantly possible (resp. certain).*

We use the threshold η in a similar way as in [20]: given a possibility distribution π on the set Ω of interpretations of a propositional logic language, $\{\omega \in \Omega \mid \pi(\omega) \geq \eta\}$ is the set of most plausible interpretations. Moreover, if ϕ is a language formula such that its necessity measure $N(\phi) = \inf\{1 - \pi(\omega) \mid \omega \models \neg\phi\}$ verifies $N(\phi) \geq \eta$ then ϕ is considered to be certain at least to the degree η.

In the following, we extract the premises justifying $\pi^*_{b(x)}(u) = \tau$ according to η. Two cases are encountered:

1. The possibility degree of "b(x) is u" is revelantly possible i.e., $\tau \geq \eta$. In this case, we rely on the relation (4.5) to establish a necessary and sufficient condition to justify $\pi^*_{b(x)}(u) = \tau$ by premises. Of course, this requires that the set J^P is non-empty, see (4.6).

2. The possibility degree of "b(x) is u" is *not* revelantly possible i.e., $\tau < \eta$. It always exists at least a premise justifying why "b(x) is u" is *not* revelantly possible.

For a given output value $u \in D_b$, let us remind that the triplets $(p_i, \mathrm{sem}_i, d_i)$ are defined in (4.3) according to (4.1) and (4.2). We select the rule premises justifiying the possibility degree $\pi^*_{b(x)}(u) = \tau$ by the following formula:

$$J_{b(x)}(u) = \begin{cases} \Big\{(p_i, \mathrm{sem}_i, d_i) \mid i \in J^P \text{ and } \theta_i = \tau\Big\} & \text{if } \tau \geq \eta. \\ \Big\{(p_i, \mathrm{sem}_i, d_i) \mid i \in \{1, 2, \ldots, n\} \text{ and } \gamma_i < \eta\Big\} & \text{if } \tau < \eta. \end{cases} \tag{4.7}$$

If $\tau \geq \eta$ and $J^P \neq \emptyset$ then, using (4.4), the equality (4.5) and the definition of $J_{b(x)}(u)$, one can check directly that we have:

Proposition 4.2 $J_{b(x)}(u) \neq \emptyset \Longleftrightarrow \pi^*_{b(x)}(u) = c_\theta$.

We prove this equivalence in two steps:

Proof 4.2

1. $J_{b(x)}(u) \neq \emptyset \Longrightarrow \pi^*_{b(x)}(u) = c_\theta$:
   As $J_{b(x)}(u) \neq \emptyset$, *there exists a triplet* $(p_{i_0}, sem_{i_0}, d_{i_0}) \in J_{b(x)}(u)$.
   Therefore, we have $t_{i_0} \leq \theta_{i_0} = \tau$, *which leads to:*

$$i_0 \in J^P \text{ and } c_\theta = \min_{i \in J^P} \theta_i \leq \theta_{i_0} = \tau = \min(c_\theta, c_t) \leq c_\theta.$$

   So, $\tau = c_\theta$.

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

By the assumption that J^P *is non empty, there exists an index* $i_0 \in J^P$ *such that:*

$$c_\theta = \min_{i \in J^P} \theta_i = \theta_{i_0}.$$

Let us show that $(p_{i_0}, sem_{i_0}, d_{i_0}) \in J_{b(x)}(u)$ *(and therefore* $J_{b(x)}(u)$ *is non-empty). We have:*

$$i_0 \in J^P, \text{ so } t_{i_0} \leq \theta_{i_0} = c_\theta.$$

By hypothesis that $\tau = c_\theta$*, so* $t_{i_0} \leq \theta_{i_0} = \tau$*. Therefore, we have:*

$$(p_{i_0}, sem_{i_0}, d_{i_0}) \in J_{b(x)}(u).$$

If $\tau \geq \eta$, $J^P \neq \emptyset$ and $J_{b(x)}(u) \neq \emptyset$, 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 $\theta_1, \theta_2, \ldots, \theta_n$.

Example 4.5 *With the hypotheses of example 4.2, let us take* $\eta = 0.1$*. By Definition (4.7), we obtain for each output attribute value:*

- $J_{fbs(x)}(low) = J_{fbs(x)}(medium) = \{(p_1, \mathsf{C}, 0.7)\}$,
- $J_{fbs(x)}(high) = \{(p_1, \mathsf{P}, 1), (p_2, \mathsf{C}, 0), (p_3, \mathsf{C}, 0)\}$.

If instead of $r_1 = 0$*, we take* $r_1 > 0.3$*, then for* $u = low$*, (4.6) holds and the corresponding set* J^P *is empty: no justification in terms of premises could be given in that case.*

4.5 Premise reduction functions

In this section, we define four functions $\mathscr{R}_\pi, \mathscr{R}_n, \mathscr{C}_\pi$ and $\mathscr{C}_n$ that reduce a compounded premise. Such premise reduction functions will be applied to the triplets of $J_{b(x)}(u)$, see (4.7) in the previous section, to form explanations of the possibility degree of an output attribute value. 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.

4.5.1 $\mathscr{P}_\pi$ and $\mathscr{P}_n$ definitions

Let $\pi : D \to [0, 1]$ be a possibility distribution on a domain D. For each subset $P \subseteq D$, we associate the following two subsets of D:

- $(P)_\pi = \{v \in P \mid \pi(v) = \Pi(P)\}$,
- $(P)_n = P \cup \{v \in \overline{P} \mid 1 - \pi(v) > N(P)\}$.

We have:

Lemma 2 $\overline{(P)_n} = \{v \in \overline{P} \mid 1 - \pi(v) = N(P)\}$.

This result is a consequence of the definition of the necessity measure:

Proof 4.3 *The definition of* $(P)_n$ *implies:*

$$\overline{(P)_n} = \overline{P} \cap \overline{\{v \in \overline{P} \mid 1 - \pi(v) > N(P)\}}.$$

We have $\{v \in \overline{P} \mid 1 - \pi(v) > N(P)\} = \overline{P} \cap \{v \in D \mid 1 - \pi(v) > N(P)\}$. *Therefore:*

$$\overline{\{v \in \overline{P} \mid 1 - \pi(v) > N(P)\}} = P \cup \{v \in D \mid 1 - \pi(v) \leq N(P)\}.$$

This leads to:

$$\begin{aligned}
\overline{(P)_n} &= \overline{P} \cap \overline{\{v \in \overline{P} \mid 1 - \pi(v) > N(P)\}} \\
&= \overline{P} \cap [P \cup \{v \in D \mid 1 - \pi(v) \leq N(P)\}] \\
&= [\overline{P} \cap P] \cup [\overline{P} \cap \{v \in D \mid 1 - \pi(v) \leq N(P)\}] \\
&= \emptyset \cup \{v \in \overline{P} \mid 1 - \pi(v) \leq N(P)\} \\
&= \{v \in \overline{P} \mid 1 - \pi(v) = N(P)\} \textit{ since } N(P) = \min_{v \in \overline{P}} (1 - \pi(v)).
\end{aligned}$$

We have:

Lemma 3 $\overline{(P)_n} = (\overline{P})_\pi$ *and* $\overline{(P)_\pi} = (\overline{P})_n$.

We prove the first relation then it will imply the second.

Proof 4.4
$\forall v \in D$, *we have:*

$$\begin{aligned}
v \in (\overline{P})_\pi &\Longleftrightarrow v \in \overline{P} \textit{ and } \pi(v) = \Pi(\overline{P}) \\
&\Longleftrightarrow v \in \overline{P} \textit{ and } 1 - \pi(v) = 1 - \Pi(\overline{P}) \\
&\Longleftrightarrow v \in \overline{P} \textit{ and } 1 - \pi(v) = N(P) \\
&\Longleftrightarrow v \in \overline{(P)_n} \textit{ thanks to Lemma 2.}
\end{aligned}$$

To prove the second relation, we substitute $\overline{P}$ to P in the first relation:

$$\overline{(\overline{P})_n} = (P)_\pi.$$

Then, we take the complementary and obtain the second relation:

$$(\overline{P})_n = \overline{(P)_\pi}.$$

Each of the relations in Lemma 3 means that the switch to the complementary $c : P \mapsto \overline{P}$ exchanges the operator $s : P \mapsto (P)_n$ with the operator $t : P \mapsto (P)_\pi$:

$$s \circ c = c \circ t \quad \text{and} \quad t \circ c = c \circ s.$$

Therefore, the two constructions $(P)_n$ and $(P)_\pi$ are equivalent:

- The first relation is exactly $(P)_n = \overline{(\overline{P})_\pi}$.
- The second relation is exactly $(P)_\pi = \overline{(\overline{P})_n}$.

Example 4.6 *In our example 4.1, let us take the possibility distribution π on $D_{cbs} = \{low, medium, high\}$ defined by:*

$$\pi(low) = 0.3, \quad \pi(medium) = 1, \quad \pi(high) = 0.$$

Given $P = \{medium, high\}$, we have:

- $(P)_\pi = \{medium\}$,
- $(P)_n = P$.

For $P' = \{low\}$, we have:

- $(P')_\pi = P'$,
- $(P')_n = \{low, high\}$.

Let us now define the two auxiliary functions $\mathscr{P}_\pi$ and $\mathscr{P}_n$.
Let a be an attribute gifted with a normalized possibility distribution π_a on its domain D_a and a proposition p of the form "$a(x) \in P$", where $P \subseteq D_a$. The normalization ensures $\max(\pi(p), \pi(\neg p)) = 1$. We introduce the following two propositions:

$$p_\pi : \text{ “}a(x) \in (P)_\pi\text{” and } p_n : \text{ “}a(x) \in (P)_n\text{”}. \tag{4.8}$$

Definition 4.2 *For the proposition p with its underlying set P and the propositions (4.8), we set:*

$$\mathscr{P}_\pi(p) = \begin{cases} p_\pi & \text{if } \pi(p) \geq \eta \\ p & \text{if } \pi(p) < \eta \end{cases} \tag{4.9}$$

$$\mathscr{P}_n(p) = \begin{cases} p_n & \text{if } n(p) \geq \eta \\ p & \text{if } n(p) < \eta \end{cases}. \tag{4.10}$$

We notice that $\mathscr{P}_\pi$ preserves the possibility degree of p i.e. $\pi(\mathscr{P}_\pi(p)) = \pi(p)$ and possibly reduces p if $\pi(p) \geq \eta$. Similarly, $\mathscr{P}_n$ preserves the necessity degree of p i.e. $n(\mathscr{P}_n(p)) = n(p)$ and reduces $\overline{P}$ if $n(p) \geq \eta$.
From Lemma 3, we deduce:

Proposition 4.3 *If $\eta \leq \pi(p) \leq 1 - \eta$ then we have:*

$$\mathscr{P}_\pi(p) = \neg\mathscr{P}_n(\neg p), \tag{4.11}$$

where $\neg p$ is the proposition formed with $\overline{P}$.

Proof 4.5 *As $\pi(p) \geq \eta$, we have $\mathscr{P}_\pi(p) = p_\pi$.
Then $\neg\mathscr{P}_\pi(p)$ is the proposition "$a(x) \in \overline{(P)_\pi}$".
As $n(\neg p) = 1 - \pi(p) \geq \eta$, then $\mathscr{P}_n(\neg p)$ is the proposition "$a(x) \in (\overline{P})_n$".
By Lemma 3, we have $\overline{(P)_\pi} = (\overline{P})_n$. Therefore, $\neg\mathscr{P}_\pi(p) = \mathscr{P}_n(\neg p)$, which is equivalent to $\mathscr{P}_\pi(p) = \neg\mathscr{P}_n(\neg p)$.*

Similarly, we also deduce from Lemma 3:

Proposition 4.4 *if $\eta \leq n(p) \leq 1 - \eta$, then we have:*

$$\mathscr{P}_n(p) = \neg\mathscr{P}_\pi(\neg p). \tag{4.12}$$

Proof 4.6 *As $n(p) \geq \eta$, we have $\mathscr{P}_n(p) = p_n$.
Then $\neg\mathscr{P}_n(p)$ is the proposition "$a(x) \in \overline{(P)_n}$".
As $\pi(\neg p) = 1 - n(p) \geq \eta$, then $\mathscr{P}_\pi(\neg p)$ is the proposition "$a(x) \in (\overline{P})_\pi$".
By Lemma 3, we have $\overline{(P)_n} = (\overline{P})_\pi$. Therefore, $\neg\mathscr{P}_n(p) = \mathscr{P}_\pi(\neg p)$, which is equivalent to $\mathscr{P}_n(p) = \neg\mathscr{P}_\pi(\neg p)$.*

Remark 4.1 *To satisfy the hypothesis of Propositions 4.3 and 4.4, we assume that* $\eta \leq 0.5$.
The normalization condition

$$\max(\pi(p), \pi(\neg p)) = \max(\pi(p), 1 - n(p)) = 1$$

implies that the two hypotheses of Propositions 4.3 and 4.4 cannot be verified at the same time.

Remark 4.2 *If* $\eta \leq \pi(p) \leq 1 - \eta$ *then the relation (4.11) holds i.e.,*

$$\mathscr{P}_\pi(p) = \neg\mathscr{P}_n(\neg p).$$

The normalization condition implies:

$$n(p) = 1 - \pi(\neg p) = 1 - 1 = 0 < \eta.$$

Therefore, we have $\mathscr{P}_n(p) = p$. *Let us compute in this case* $\neg\mathscr{P}_\pi(\neg p)$.
We have $\pi(\neg p) = 1 \geq \eta$ *then* $\mathscr{P}_\pi(\neg p)$ *is the proposition "*$a(x) \in (\overline{P})_\pi$*" from which we deduce that* $\neg\mathscr{P}_\pi(\neg p)$ *is "*$a(x) \in \overline{(\overline{P})_\pi}$*" i.e., "*$a(x) \in (P)_n$*".*
As $(P)_n = P \cup \{v \in \overline{P} \mid 1 - \pi_{a(x)}(v) > N(P)\}$, *we may have* $(P)_n \neq P$ *and then* $\mathscr{P}_n(p) \neq \neg\mathscr{P}_\pi(\neg p)$. *Generally, we cannot have the relation (4.11) and the relation (4.12) at the same time.*

Example 4.7 *In our example 4.2, the proposition "act(x) ∈ {dinner, drink-coffee, lunch}" is reduced by* $\mathscr{P}_\pi$ *to "act(x) ∈ {drink-coffee}" while* $\mathscr{P}_n$ *keeps it as is.*

$\mathscr{P}_\pi$ and $\mathscr{P}_n$ are used in the definitions of the premise reduction functions. We continue by defining subsets of propositions of a compounded premise, which are also used in the definitions of the premise reduction functions.

4.5.2 Subsets of propositions of a compounded premise

In the following, let $p = p_1 \wedge p_2 \wedge \cdots \wedge p_k$ be a compounded premise, where p_j for $j = 1, 2, \ldots, k$, is a proposition of the form "$a_j(x) \in P_j$" with $P_j \subseteq D_{a_j}$. When p is not considered as relevantly possible i.e., $\pi(p) < \eta$, we introduce the following two sets of propositions extracted from p with respect to the threshold η:

- a set of propositions that are relevantly possible:

$$A_p^\pi = \Big\{p_j \mid \pi(p_j) \geq \eta \text{ for } j = 1, \ldots, k\Big\}. \quad (4.13)$$

- a set of propositions that are not considered relevantly possible:

$$B_p^\pi = \Big\{p_j \mid \pi(p_j) < \eta \text{ for } j = 1, \ldots, k\Big\}. \quad (4.14)$$

Similarly, when p is not considered as relevantly certain i.e., $n(p) < \eta$, we introduce the following two sets of propositions extracted from p with respect to the threshold η:

- a set of propositions that are relevantly certain:

$$A_p^n = \Big\{ p_j \mid n(p_j) \geq \eta \text{ for } j = 1, \ldots, k \Big\}. \tag{4.15}$$

- a set of propositions that are not considered relevantly certain:

$$B_p^n = \Big\{ p_j \mid n(p_j) < \eta \text{ for } j = 1, \ldots, k \Big\}. \tag{4.16}$$

These four sets of propositions are used in the definitions of the reduction functions, which we now introduce.

4.5.3 Extracting justifications: $\mathscr{R}_\pi$ function

Given the compounded premise $p = p_1 \wedge p_2 \wedge \cdots \wedge p_k$,the function $\mathscr{R}_\pi$ returns the structure responsible for $\pi(p)$. If p is relevantly possible with respect to the threshold η, $\mathscr{R}_\pi$ returns the conjunction of the propositions $\mathscr{P}_\pi(p_j)$. Otherwise, if p is not considered relevantly possible, $\mathscr{R}_\pi$ returns the conjunction of the propositions in B_p^π, see (4.14). The reduction function $\mathscr{R}_\pi$ extends $\mathscr{P}_\pi$ in the following sense:

$$\mathscr{R}_\pi(p) = \begin{cases} \bigwedge_{j=1}^{k} \mathscr{P}_\pi(p_j) & \text{if } \pi(p) \geq \eta \\ \bigwedge_{p_j \in B_p^\pi} p_j & \text{if } \pi(p) < \eta \end{cases}. \tag{4.17}$$

We note that $\forall p_j \in B_p^\pi \; \mathscr{P}_\pi(p_j) = p_j$. In the two cases i.e. $\pi(p) \geq \eta$ and $\pi(p) < \eta$, we have $\pi(\mathscr{R}_\pi(p)) = \pi(p)$: the function $\mathscr{R}_\pi$ preserves the possibility degree of p.

Example 4.8 *In our example 4.2:*

- *for the premise* p_1 *of the rule* R^1*,* $\mathscr{R}_\pi$ *returns the conjunction of the propositions "act(x) ∈ {drink-coffee}" and "cbs(x) ∈ {medium}",*

- *for the premise* p_2 *of* R^2*,* $\mathscr{R}_\pi$ *returns the proposition "act(x) ∈ {long-sleep, sport, walking}" and*

- *for the premise* p_3 *of* R^3*,* $\mathscr{R}_\pi$ *returns the proposition "act(x) ∈ {alcohol-consumption, breakfast}".*

4.5.4 Extracting justifications: $\mathscr{R}_n$ function

Similarly, the reduction function $\mathscr{R}_n$ returns the structure responsible for the necessity degree $n(p)$ of p, which is the conjunction of the propositions $\mathscr{P}_n(p_j)$ that make p relevantly certain or not. In particular, if p is not considered relevantly certain, $\mathscr{R}_n$ returns the conjunction of the propositions in B_p^n, see (4.16). The reduction function $\mathscr{R}_n$ extends $\mathscr{P}_n$ in the following sense:

$$\mathscr{R}_n(p) = \begin{cases} \bigwedge_{j=1}^{k} \mathscr{P}_n(p_j) & \text{if } n(p) \geq \eta \\ \bigwedge_{p_j \in B_p^n} p_j & \text{if } n(p) < \eta \end{cases}. \tag{4.18}$$

We note that $\forall p_j \in B_p^n$, $\mathscr{P}_n(p_j) = p_j$. $\mathscr{R}_n$ preserves the necessity degree of p i.e., $n(\mathscr{R}_n(p)) = n(p)$.

Example 4.9 *In our example 4.2:*

- *for the premise p_1 of the rule R^1, $\mathscr{R}_n$ returns p_1 as it is,*
- *for the premise p_2 of R^2, $\mathscr{R}_n$ returns the proposition "act(x) $\in$ {long-sleep, sport, walking}" and*
- *for the premise p_3 of R^3, $\mathscr{R}_n$ returns the proposition "act(x) $\in$ {alcohol-consumption, breakfast}".*

4.5.5 Extracting unexpectedness: $\mathscr{C}_\pi$ function

Intuitively, with respect to the threshold η, if p is not relevantly possible i.e, $\pi(p) < \eta$, $\mathscr{C}_\pi$ returns a conjunction of propositions, called an unexpectedness, which is *not* involved in the determination of $\pi(p)$ although relevantly possible. To use $\mathscr{C}_\pi$, the set of relevantly possible propositions of p must be non-empty. In other words, we suppose $A_p^\pi \neq \emptyset$, see (4.13).
The reduction function $\mathscr{C}_\pi$ returns the conjunction of the propositions $\mathscr{P}_\pi(p_j)$ such that $\pi(p_j) \geq \eta$:

$$\mathscr{C}_\pi(p) = \bigwedge_{p_j \in A_p^\pi} \mathscr{P}_\pi(p_j). \tag{4.19}$$

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 $\pi(p_j)$.

Example 4.10 *In our example 4.2, for the premises of R^2 and R^3, $\mathscr{C}_\pi$ returns for both "cbs(x) $\in$ {medium}".*
In other words, although it may be otherwise, the fact that the current blood glucose level is medium is not involved in determining the possibility degree of the premise of R^2 and that of the premise of R^3.

4.5.6 Extracting unexpectedness: $\mathscr{C}_n$ function

Similarly, if p is not relevantly certain i.e., $n(p) < \eta$, $\mathscr{C}_n$ returns a conjunction of propositions, called an unexpectedness, which is *not* involved in the determination of $n(p)$ although relevantly certain. Analogously, in order to define $\mathscr{C}_n(p)$ we suppose $A_p^n \neq \emptyset$, see (4.15), i.e., the set of relevantly certain propositions of p is non-empty.
The function $\mathscr{C}_n$ returns the conjunction of the propositions $\mathscr{P}_n(p_j)$ such that $n(p_j) \geq \eta$:

$$\mathscr{C}_n(p) = \bigwedge_{p_j \in A_p^n} \mathscr{P}_n(p_j). \tag{4.20}$$

We notice that 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 $n(p_j)$.

Example 4.11 *In our example 4.2, for the premises of R^2 and R^3, $\mathscr{C}_n$ returns for both "cbs(x) $\in$ {low, medium}".*

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

We remind that the triplets of $J_{b(x)}(u)$ associated to $\pi^*_{b(x)}(u)$, see Definition (4.7), are of the form (p, sem, d), see Notation 1. In order to form a justification of $\pi^*_{b(x)}(u)$ and to extract the unexpectedness of $\pi^*_{b(x)}(u)$, we will apply the premise reduction functions to the triplets of $J_{b(x)}(u)$.
To apply in an appropriate way the reduction functions $\mathscr{R}_\pi$ and $\mathscr{R}_n$ to a triplet (p, sem, d), we introduce the function $\mathscr{S}_{\mathscr{R}}$:

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

Similarly, to apply $\mathscr{C}_\pi$ and $\mathscr{C}_n$, we introduce the function $\mathscr{S}_{\mathscr{C}}$, which relies on the set of propositions A_p^π and A_p^n, see (4.13) and (4.15):

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

The justification of $\pi^*_{b(x)}(u)$ is formed by applying $\mathscr{S}_{\mathscr{R}}$ to the triplets of $J_{b(x)}(u)$, see (4.7):

$$\text{Justification}_{b(x)}(u) = \{\mathscr{S}_{\mathscr{R}}(p, sem, d) \mid (p, sem, d) \in J_{b(x)}(u)\}. \tag{4.21}$$

The possibilistic expressions in the triplets of (4.21) are sufficient to justify "$b(x)$ is u at a possibility degree $\pi^*_{b(x)}(u)$".

Example 4.12 *In our example 4.2, we form the justification for each output attribute value and give, for each, an interpretation:*

- $Justification_{fbs(x)}(low) = Justification_{fbs(x)}(medium) = \{(\mathscr{R}_n(p_1), \mathsf{C}, 0.7)\} = \{(p_1, \mathsf{C}, 0.7)\}$, *which could be interpreted as:*

  > *"A future low (resp. medium) blood sugar level is evaluated as not very possible. This is mainly due to the fact that it is quite certain that the activity consists of drinking coffee, lunch or dinner and that the current blood sugar level is medium or high".*

- $Justification_{fbs(x)}(high) = \{(\mathscr{R}_\pi(p_1), \mathsf{P}, 1), (\mathscr{R}_n(p_2), \mathsf{C}, 0), (\mathscr{R}_n(p_3), \mathsf{C}, 0)\}$. *A natural language explanation could be:*

  > *"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.".*

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:

$$\text{Unexpectedness}_{b(x)}(u) = \{\mathscr{S}_\mathscr{C}(p, sem, d) \mid (p, sem, d) \in J_{b(x)}(u)\}. \quad (4.22)$$

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 [39].

Remark 4.3 *It is clear that any triplet of $J_{b(x)}(u)$ is in the domain of the function $\mathscr{S}_\mathscr{R}$. It is not the same for the function $\mathscr{S}_\mathscr{C}$: not all triplets of $J_{b(x)}(u)$ are necessarily in the domain of the function $\mathscr{S}_\mathscr{C}$. For example, in the case where $\tau < \eta$, a triplet $(p_i, sem_i, d_i) \in J_{b(x)}(u)$ of the form:*

$$(p_i, sem_i, d_i) = (p_i, \mathsf{C}, n(p_i))$$

is not in the domain of the function $\mathscr{S}_\mathscr{C}$. Indeed, in this case, we have $d_i > \eta$ since:

$$\begin{aligned} d_i = 1 - \rho_i &\geq 1 - \beta_i \\ &= 1 - \gamma_i \\ &> 1 - \eta \geq \eta \quad \text{if} \quad \eta \leq 0.5. \end{aligned}$$

Example 4.13 *For our example 4.2, we form the unexpectedness of the possibility degree of each output attribute value. An unexpectedness exists for* $\pi^*_{fbs(x)}(high)$*. No unexpectedness exists for both* $\pi^*_{fbs(x)}(low)$ *and* $\pi^*_{fbs(x)}(medium)$*. As* $\mathscr{C}_n(p_2) = \mathscr{C}_n(p_3)$*, we obtain* $Unexpectedness_{fbs(x)}(high) = \{(\mathscr{C}_n(p_2), \mathsf{C}, 1)\}$*. The interpretation is as follows:*

> *"It is possible for the patient's blood sugar level to become high even if it is certain that the current blood sugar level is low or medium".*

4.7 Example

In the following, we propose another example of a possibilistic rule-based system. It determines the insulin dose required by a patient according to some factors [27].

4.7.1 Rule base

We consider nine input attributes: *planned-foods (pf)*, *planned-alcohol (pa)*, *current-bloodsugar (cbs)*, *planned-physicalactivity (ppa)*, *planned-sleep (ps)*, *water-intake (wi)*, *last-hypoglycemia (lh)*, *previous-slept-duration (psd)* and *environmental-temperature (et)*. We have D_{pf} = {*no, standard-meal, high-fat-foods*}, D_{pa} = {*no, low, important*}, D_{cbs} = {*low, medium, high*}, D_{ppa} = {*no, short, long*}, D_{ps} = {*no, short, long, very-long*}, D_{wi} = {*sufficient, insufficient*}, D_{lh} = {*no, long-time-ago, recent, very-recent*}, D_{psd} = {*very-short, short, long, very-long*} and D_{et} = {*cold, warm, hot*}. The output attribute is named *insulin-dose (id)* such that D_{id} = {*low, medium, high*}.
The rule base is composed of five rules (Table 4.2).

	pf	pa	cbs	ppa	ps	wi	lh	psd	et	id
R^1	standard-meal, high-fat-foods	no, low								medium, high
R^2			high	no, short						low
R^3	no		low, medium		long, very-long					low
R^4						sufficient	no, long-time-ago	long, very-long		low, medium
R^5									cold, warm	low, medium

Table 4.2: Rule base for determining an insulin dose.

We arbitrarily take the following rule parameters $s_1 = 0.3, s_2 = 1, s_3 = 0.7, s_4 = 0.4, s_5 = r_1 = r_2 = r_3 = r_4 = 0$ and $r_5 = 1$.

4.7.2 Inference

We assume that $\pi_{pf(x)}$: $\langle$no : 0, standard-meal : 1, high-fat-foods : 0.2$\rangle$, $\pi_{pa(x)}$: $\langle$no : 1, low : 0, important : 0$\rangle$, $\pi_{cbs(x)}$: $\langle$low : 0.3, medium : 1, high : 0$\rangle$, $\pi_{ppa(x)}$: $\langle$no : 1, short : 0, long : 0$\rangle$, $\pi_{ps(x)}$: $\langle$no : 1, short : 0, long : 0, very-long : 0$\rangle$, $\pi_{wi(x)}$: $\langle$sufficient : 1, insufficient : 0.4$\rangle$, $\pi_{lh(x)}$: $\langle$no : 0.3, long-time-ago : 1, recent : 0, very-recent : 0$\rangle$, $\pi_{psd(x)}$: $\langle$: very-short : 0, short : 1, long : 0.2, very-long : 0$\rangle$ and $\pi_{et(x)}$: $\langle$cold : 0, warm : 1, hot : 0.2$\rangle$.
For the premises of the rules, we get $\lambda_1 = 1, \rho_1 = 0, \lambda_2 = 0, \rho_2 = 1, \lambda_3 = 0, \rho_3 = 1, \lambda_4 = 0.2, \rho_4 = 1, \lambda_5 = 1,$ and $\rho_5 = 0.2$. Therefore:

- $\pi^*_{\text{id(x)}}(\mathit{low}) = \min(\beta_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = \min(0, 1, 0.7, 0.4, 1) = 0$,
- $\pi^*_{\text{id(x)}}(\mathit{medium}) = \min(\alpha_1, \beta_2, \beta_3, \alpha_4, \alpha_5) = \min(1, 1, 1, 0.4, 1) = 0.4$ and
- $\pi^*_{\text{id(x)}}(\mathit{high}) = \min(\alpha_1, \beta_2, \beta_3, \beta_4, \beta_5) = \min(1, 1, 1, 1, 1) = 1$,

as $\alpha_1 = \max(s_1, \lambda_1) = \max(0.3, 1) = 1$, $\beta_1 = \max(r_1, \rho_1) = \max(0, 0) = 0$, $\alpha_2 = \max(s_2, \lambda_2) = \max(1, 0) = 1$, $\beta_2 = \max(r_2, \rho_2) = \max(0, 1) = 1$, $\alpha_3 = \max(s_3, \lambda_3) = \max(0.7, 0) = 0.7$, $\beta_3 = \max(r_3, \rho_3) = \max(0, 1) = 1$ $\alpha_4 = \max(s_4, \lambda_4) = \max(0.4, 0.2) = 0.4$, $\beta_4 = \max(r_4, \rho_4) = \max(0, 1) = 1$, $\alpha_5 = \max(s_5, \lambda_5) = \max(0, 1) = 1$ and $\beta_5 = \max(r_5, \rho_5) = \max(1, 0.2) = 1$.

4.7.3 Extraction of premises

Using the relation (4.4), we form the following sets associated to each output attribute value:

- $J^P_l = \{1, 5\}$ and $J^R_l = \{1, 2, 3, 4\}$ for low,
- $J^P_m = \{1, 2, 3, 5\}$ and $J^R_m = \{4\}$ for medium,
- $J^P_h = \{1, 2, 3, 4\}$ and $J^R_h = \{5\}$ for high.

We deduce:

- $c^l_\theta = c^l_t = 0 = \pi_{\text{id(x)}}(\text{low})$.
- $c^m_\theta = 1$ and $c^m_t = 0.4 = \pi_{\text{id(x)}}(\text{medium})$.
- $c^h_\theta = c^h_t = 1 = \pi_{\text{id(x)}}(\text{high})$.

By Definition 4.7, we obtain for each output attribute value:

- $J_{\text{id(x)}}(\text{low}) = \{(p_1, \mathsf{C}, 1)\}$,
- $J_{\text{id(x)}}(\text{medium}) = \emptyset$,
- $J_{\text{id(x)}}(\text{high}) = \{(p_1, \mathsf{P}, 1), (p_2, \mathsf{C}, 0), (p_3, \mathsf{C}, 0), (p_4, \mathsf{C}, 0)\}$.

We cannot form a justification or an unexpectedness for the output attribute value *medium* because $J_{\text{id(x)}}(\text{medium})$ is empty.

4.7.4 Application of premise reduction functions

In order to form the justification of the output attribute value *low* and that of *high*, we apply $\mathscr{R}_\pi$ and $\mathscr{R}_n$ to the selected premises in $J_{\mathrm{id(x)}}(\mathrm{low})$ and $J_{\mathrm{id(x)}}(\mathrm{high})$, respectively:

- For the premise p_1 of R^1, $\mathscr{R}_\pi$ returns the conjunction of the propositions "pf(x) $\in$ {standard-meal}" and "pa(x) $\in$ {no}" while $\mathscr{R}_n$ returns p_1 as it is.
- For the premise of R^2, $\mathscr{R}_n$ returns the proposition "cbs(x) $\in$ {high}".
- For the premise of R^3, $\mathscr{R}_n$ returns the conjunction of the propositions "pf(x) $\in$ {no}" and "ps(x) $\in$ {long,very-long}".
- When we apply $\mathscr{R}_n$ to the premise of R^4, we obtain the proposition "psd(x) $\in$ {long, very-long}".

To obtain the unexpectedness for *high*, we apply $\mathscr{C}_n$:

- For the premise of R^2 $\mathscr{C}_n$ returns a proposition: "ppa(x) $\in$ {no, short}".
- For the premise of R^3, $\mathscr{C}_n$ returns "cbs(x) $\in$ {low,medium}".
- For the premise of R^4, $\mathscr{C}_n$ returns the conjunction of "lh(x) $\in$ {no, long-time-ago}" and "wi(x) $\in$ {sufficient}".

The other output attribute values have not an unexpectedness.

4.7.5 Justifications

We form the justification of *low* and *high*:

- $\mathrm{Justification}_{\mathrm{id(x)}}(\mathrm{low}) = \{(\mathscr{R}_n(p_1), \mathsf{C}, 1)\} = \{(p_1, \mathsf{C}, 1)\}$, which could be interpreted as:

  > "*It is not possible to provide a small dose of insulin to the patient. This is mainly due to the fact that it is certain that he wants a standard meal or high-fat foods. It is also certain that he planned little or no alcohol*".

- $\mathrm{Justification}_{\mathrm{id(x)}}(\mathrm{high}) = \{(\mathscr{R}_\pi(p_1), \mathsf{P}, 1),\ (\mathscr{R}_n(p_2), \mathsf{C}, 0),\ (\mathscr{R}_n(p_3), \mathsf{C}, 0),$ $(\mathscr{R}_n(p_4), \mathsf{C}, 0)\}$. A natural language explanation could be:

  > "*It is possible to administer a high dose of insulin to the patient. In fact, he planned to have a standard meal and no alcohol. In addition, the following facts are assessed as not certain: his current blood sugar level is high, he doesn't plan to eat, he wants to sleep a lot.*".

4.7.6 Unexpectedness

We obtain the unexpectedness for *high*:

$$\text{Unexpectedness}_{\text{id(x)}}(\text{high}) = \{(\mathscr{C}_n(p_2), \mathsf{C}, 1), (\mathscr{C}_n(p_3), \mathsf{C}, 1), (\mathscr{C}_n(p_4), \mathsf{C}, 0.6)\}.$$

The interpretation is as follows:

> "*It is possible to administer a high dose of insulin to the patient even though it is certain that he has planned to do little or no physical activity, the current blood sugar level is low or medium, his last hypoglycemia was a long time ago (or he did not have any) and he drank enough water*".

4.8 Conclusion

In this chapter, we studied 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 premises the possibility degree of each output attribute value. We then performed reductions on the selected premises, in order to form two kind of explanations: the justification and the unexpectedness of the possibility degree of an output attribute value.

An evaluation of our explanations would be useful in order to know if they are appropriate to the users' needs and when they are expected by the user, especially for the unexpectedness. From each of the selected premise, our method picks some of its propositions, and reduces or extends them. We should check if this allows a better understanding of the decisions. We may also compare our explanations with others, which would be generated by performing other operations on the selected premises e.g., letting the propositions be as they are. To go further in the evaluation process of our method, it would be interesting to measure to what extent the parameter η has an impact on the understanding of the decisions.

Finally, as the explanations are the outcomes of numerous analytical operations, and in order to facilitate the inspection of their content, they should be represented in terms of graphs. In Part D, a representation of explanations of possibilistic inference decisions is developed. It is based on the justification and the unexpectedness of the possibility degree of an output attribute.

In the next chapter, similarly to the possibilistic case, we introduce methods for extracting justifications and unexpectedness of inference results of a fuzzy rule-based system composed of possibility rules.

Chapter 5

Justification of fuzzy rule-based system inference results

The work in this chapter is based on a paper published in a conference: Baaj, I., & Poli, J. P. (2019, June). Natural language generation of explanations of fuzzy inference decisions. In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.

In this chapter, we elaborate on the explanatory capabilities of a fuzzy rule-based system composed of possibility rules (Mamdani fuzzy inference system), where the premises of the rules are conjunctions of fuzzy propositions. We focus on explaining semantically the inferred conclusions of a Mamdani system, without considering the use of a defuzzification process [102].
For this purpose, in Section 5.1, we set the notations used for the main objects of an inference of a Mamdani system: activation degree of a rule, inferred conclusion and the possibility distribution of the variable of the rule conclusions. Then, we state in Proposition 5.1 the main result: the total inferred conclusion satisfies the semantics α^*-possible in the sense of Dubois-Prade [49], where α^* is the maximum of the activation degrees of the rules. This result is proved in two steps (Subsections 5.1.1 and 5.1.2). In Section 5.2, we begin by adding some useful notations for justifying the inference results of a Mamdani system. We introduce an example of such a system, which we will use to illustrate all the constructions that follow. Then, we justify by a relevant subset of rule premises each of the inferred conclusions of any Mamdani system. In Section 5.3, we introduce two premise reduction functions, which are similar to those introduced for the possibilistic case (Chapter 4). By applying them to the premises selected for justifying an inferred conclusion, such functions allow us to form two types of explanations (Section 5.4):

- The *justification* of a conclusion, which is a set of fuzzy logic expressions (conjunctions of fuzzy propositions) sufficient to justify, semantically, the inferred conclusion. It is formed by applying a reduction function to the selected premises. Such reduction function is based on our previous work, see [15].

- The *unexpectedness* of a conclusion, which is a set of fuzzy logic expressions extracted by applying another premise reduction function to the selected premises. Such fuzzy logic expressions are related to the considered conclusion in the following sense: although there may appear to be a potential incompatibility between each of the fuzzy logic expressions and the considered conclusion, they are not involved in the determination of the inferred conclusion.

Such explanations are formulated by performing treatments on the rule premises, which are conjunctions of fuzzy propositions. In Part D, such explanations will be graphically represented by conceptual graphs. For this purpose, we will show that we can naturally represent conjunctions of fuzzy propositions by conceptual graphs. Note that such graphs do not allow representing logical disjunctions, and only adopt a limited form of negation (called atomic negation) [34].
Finally, we conclude with some perspectives (Section 5.5).

5.1 Semantics of the total inferred conclusion of a Mamdani rule-based system

In this section, we study the semantics of the inferred conclusions of a Mamdani fuzzy inference system. Let $R^1, R^2, \ldots, R^n$ be the n rules of a Mamdani system and u^0 be a crisp input for the variables that appear in the n rules. For each $i = 1, 2, \ldots, n$, we set:

- $R^i = (p_i, c_i)$, where the premise $p_i = p_1^i \wedge p_2^i \wedge \cdots \wedge p_{k_i}^i$ is a conjunction of fuzzy propositions and $c_i = (Z, O_i)$ is a fuzzy proposition with the same variable Z of a linguistic variable $z = (Z, V, T_z)$ and a term $O_i \in T_z$. Each $p_j^i = (X_j^i, A_j^i)$ is composed of the variable X_j^i of a linguistic variable a_j^i, and a term $A_j^i \in T_{a_j^i}$. Such proposition models the statement "X_j^i is A_j^i is possible", using the (guaranteed) possibility distribution $\delta_{X_j^i} = \mu_{A_j^i}$.

- α_{p_i} is the fuzzy degree of the premise p_i (or the activation degree of the rule R^i) deduced from the crisp input u^0.

- O_i^* is the inferred fuzzy set of the conclusion that we get by firing the rule R^i.

- $\delta_{Z,i}$ is the (guaranteed) possibility distribution associated to the proposition "Z is O_i^*" of the rule R^i [67, 100].

Then, the main objects associated by the inference to the input u^0 are:

- $\alpha^* = \max_i \alpha_{p_i}$,
- $O^* = \cup_i O_i^*$ is the total inferred fuzzy set that we get by firing the n rules. Such fuzzy set is sometimes called "Mamdani distribution" [76].
- $\delta_Z = \max_i \delta_{Z,i}$ is a natural (guaranteed) possibility distribution for Z associated to the proposition "Z is O^*" [67, 100].

With these notations, we claim:

Proposition 5.1 *Z is O^* is α^*-possible in the sense of Dubois-Prade [49].*

This result is already proven in [49] for a single rule $R = (p, c)$ where $p = (X, A)$ is a fuzzy proposition. For the proof of the general case, we first extend Dubois-Prade's result for a single rule $R = (p, c)$ where the premise p is a conjunction of k fuzzy propositions, then we deduce the general case of a system of n rules.

5.1.1 Single rule with a compounded premise

We remind that a Mamdani rule is a possibility rule, where the t-norm is the min function (Chapter 1). Let $R = (p, c)$ be a possibility rule such that:

$$p = X_1 \text{ is } A_1 \wedge X_2 \text{ is } A_2 \wedge \cdots \wedge X_k \text{ is } A_k, \tag{5.1}$$

is a conjunction of k fuzzy propositions (X_j, A_j) defined by the linguistic variables $a_j = (X_j, U_j, T_{a_j})$ and the terms $A_j \in T_{a_j}$. The conclusion c is a fuzzy proposition $c = (Z, O)$ defined by a linguistic variable $z = (Z, V, T_z)$ and $O \in T_z$. Let u^0 be a crisp input such that:

$$u^0 = (u_1^0, u_2^0, \ldots, u_k^0) \in U_1 \times U_2 \times \cdots \times U_k.$$

The Mamdani inferred conclusion associated to these data is the fuzzy subset O^* of V whose membership function μ_{O^*} is defined by:

$$\forall v \in V, \mu_{O^*}(v) = \min(\mu_O(v), \alpha_p) \text{ where } \alpha_p = \min_j(\mu_{A_j}(u_j^0)). \tag{5.2}$$

Let δ_Z be the possibility distribution of the conclusion variable Z of the rule R (relation 1.13). Using the guaranteed possibility measure Δ associated with δ_Z evaluated with the fuzzy set O^*, see (1.5), we claim:

Proposition 5.2 *Z is O^* is $\alpha_p -$ possible.*

We prove the proposition 5.2 as follows:

Proof 5.1 *We must prove that* $\Delta(O^*) \geq \alpha_p$*, which is equivalent to:*

$$\forall v \in V, \mu_{O^*}(v) \rightarrow_g \delta_Z(v) \geq \alpha_p.$$

Using the equivalence taken from [49]:

$$a \rightarrow b \geq c \Longleftrightarrow b \geq \min(a, c),$$

It is equivalent to prove:

$$\forall v \in V, \delta_Z(v) \geq \min(\mu_{O^*}(v), \alpha_p).$$

As for any $v \in V$*, we have* $\mu_{O^*}(v) = \min(\mu_O(v), \alpha_p) = \min(\mu_{O^*}(v), \alpha_p)$*, it is sufficient to prove:*

$$\forall v \in V, \delta_Z(v) \geq \min(\mu_O(v), \alpha_p).$$

We have:

$$\begin{aligned}\delta_Z(v) = & \sup_{(u_1,u_2,\ldots,u_k)\in U_1\times U_2\times\cdots\times U_k} \min(\min_j \mu_{A_j}(u_j), \mu_O(v)) \\ & \geq \min(\min_j \mu_{A_j}(u_j^0), \mu_O(v)) = \min(\alpha_p, \mu_O(v)).\end{aligned}$$

We have proven:

$$\forall v \in V, \delta_Z(v) \geq \mu_{O^*}(v) = \min(\alpha_p, \mu_O(v)). \tag{5.3}$$

The inequality (5.3) means that δ_Z is a guaranteed possibility distribution for the proposition "Z is O^*" [67, 100].

5.1.2 Mamdani fuzzy inference system of n rules

For the case of n rules, we adopt the notations of the beginning of the section. We must prove that $\Delta(O^*) \geq \alpha^*$:

Proof 5.2 *To prove the inequality* $\Delta(O^*) \geq \alpha^*$ *is equivalent to prove that:*

$$\forall v \in V, \mu_{O^*}(v) \rightarrow_g \delta_Z(v) \geq \alpha^*.$$

Using the equivalence:

$$a \rightarrow b \geq c \Longleftrightarrow b \geq \min(a, c),$$

it is equivalent to prove:

$$\forall v \in V, \delta_Z(v) \geq \min(\mu_{O^*}(v), \alpha^*).$$

As for any $v \in V$, we have $\mu_{O^*}(v) = \max_i(\min(\mu_{O_i}(v), \alpha_{p_i}))$, we deduce:

$$\forall v \in V, \mu_{O^*}(v) \leq \alpha^*.$$

The possibility distribution δ_Z is the disjunctive combination of the (guaranteed) possibility distributions $\delta_{Z,i}$ for $i = 1, 2, \cdots, n$, where each one is associated to the proposition "Z is O_i^*" of R^i:

$$\delta_Z(v) = \max_i \delta_{Z,i}(v).$$

For each $i = 1, \ldots, n$, we have:

$$\forall v \in V, \delta_{Z,i}(v) \geq \min(\mu_{O_i}(v), \alpha_{p_i}) = \mu_{O_i^*}(v).$$

We deduce:

$$\forall v \in V, \delta_Z(v) = \max_i \delta_{Z,i}(v) \geq \max(\mu_{O_i^*}(v)) = \mu_{O^*}(v) = \min(\mu_{O^*}(v), \alpha^*). \quad (5.4)$$

Therefore, (5.4) implies that δ_Z is a guaranteed possibility distribution for the proposition "Z is O^*" [67, 100].

5.2 Justifying an inferred conclusion of a Mamdani fuzzy inference system

In this section, for the study of explainability of a Mamdani fuzzy inference system, we slightly change some previous notations and add some new ones. We give an example of a Mamdani system, which will be used to illustrate all the constructions. Then, we introduce a method for justifying each inferred conclusions of any Mamdani system by a relevant subset of rule premises.

We denote by $O_1, O_2, \ldots, O_k$ the *distinct* linguistic terms which appear in the n conclusions of the system i.e., $O_i \neq O_j$ for $i \neq j$. For $j = 1, 2, \ldots, k$, let $B_{(Z,O_j)}$ be the set of rules whose conclusion is (Z, O_j):

$$B_{(Z,O_j)} = \{R^i \mid R^i = (p_i, c_i) \text{ such that } c_i = (Z, O_j)\}.$$

For the sake of simplicity, we denote by α_e the fuzzy degree of any fuzzy logic expression e. By definition, we have $\alpha_{R^i} = \alpha_{p_i}$. Finally, for $j = 1, 2, \ldots, k$

we denote by $\alpha^*_{(Z,O_j)}$ the highest activation degree among all the activation degrees of the fuzzy rules in $B_{(Z,O_j)}$:

$$\alpha^*_{(Z,O_j)} = \max_{R^i \in B_{(Z,O_j)}} \alpha_{R^i}. \tag{5.5}$$

For the interpretation and considering that we have described a Mamdani system in terms of possibility rules in the previous section, one can notice that the statements which are modeled by any fuzzy proposition in a premise or a conclusion use the same semantics: *possible*, which we note P. From this, one can represent the interpretation of the premise p of a possibility rule $R = (p, c)$ by the following triplet:

Notation 2 *Let* $p = X_1$ *is* $A_1 \wedge X_2$ *is* $A_2 \wedge \cdots \wedge X_k$ *is* A_k *be a premise of a possibility rule* $R = (p, c)$ *whose (guaranteed) joint possibility distribution is* $\delta_{X_1,X_2,\cdots,X_k}$ *(relation 1.11) and a crisp input* $u^0 = (u^0_1, u^0_2, \ldots, u^0_k)$*. The triplet* (p, sem, d) *denotes* $(p, \mathsf{P}, \alpha_p)$*, where* $sem = \mathsf{P}$ *is the semantics attached to the fuzzy degree* $\alpha_p = \delta_{X_1,X_2,\ldots,X_k}(u^0_1, u^0_2, \ldots, u^0_k) = \min_j(\mu_{A_j}(u^0_j))$*.*

A triplet (p, sem, d) may be used to interpret the premise of other fuzzy rules such as certainty rules [100] but with an adapted semantics and degree.
Our notations are illustrated by the following example of a fuzzy rule-based system:

Example 5.1 *We consider a Mamdani fuzzy inference system that controls the blood glucose level of a patient with type 1 diabetes. The prediction is based on some factors identified by [27] that may affect the blood glucose level. The rule base is composed of thirtheen rules built from five input linguistic variables: activity (Act), current-bloodsugar (Cbs), last-hypoglycemia (Lh) and sleep-quality (Sq), water-intake (Wi) and one output linguistic variable: future-bloodsugar (Fbs). For simplicity, we describe in the following the linguistic terms associated to each linguistic variable and denote the name of the variable by its associated linguistic variable. Nine rules are used to predict the future blood sugar level (Low, Medium, High) according to the chosen activity (Eat, Sport, AlcoholConsumption) and the current blood sugar level (Low, Medium, High):*

- R^1*: If Act is Eat and Cbs is Low then Fbs is Medium,*
- R^2*: If Act is Eat and Cbs is Medium then Fbs is High,*
- R^3*: If Act is Eat and Cbs is High then Fbs is High,*
- R^4*: If Act is Sport and Cbs is Low then Fbs is Low,*
- R^5*: If Act is Sport and Cbs is Medium then Fbs is Low,*

- *R^6: If Act is Sport and Cbs is High then Fbs is Medium,*
- *R^7: If Act is AlcoholConsumption and Cbs is Low then Fbs is Low,*
- *R^8: If Act is AlcoholConsumption and Cbs is Medium then Fbs is Low,*
- *R^9: If Act is AlcoholConsumption and Cbs is High then Fbs is Medium.*

Two rules help to predict a future lower blood sugar level if the last known hypoglycemia is recent:

- *R^{10}: If Lh is Recent and Cbs is Medium then Fbs is Low,*
- *R^{11}: If Lh is Recent and Cbs is High then Fbs is Medium.*

Otherwise, if the last hypoglycemia occurred a long time ago, this would not influence the prediction. In fact, if the patient has experienced a hypoglycemia within twelve hours, it is possible that he or she will experience another one [27, 38].
Two other rules check respectively if the patient has not slept enough or drunk enough water. In both cases, the blood sugar level becomes high [27]:

- *R^{12}: If Sq is Bad then Fbs is High,*
- *R^{13}: If Wi is Insufficient then Fbs is High.*

The conclusions of the rule base are (Fbs, Low), $(Fbs, Medium)$ and $(Fbs, High)$. We have $B_{(Fbs,Low)} = \{R^4, R^5, R^7, R^8, R^{10}\}$, $B_{(Fbs,Medium)} = \{R^1, R^6, R^9, R^{11}\}$ and $B_{(Fbs,High)} = \{R^2, R^3, R^{12}, R^{13}\}$.
The patient's inputs are i_{Act}, i_{Cbs}, i_{Lh}, i_{Sq} and i_{Wi}. In our example, the patient wants to eat: $\mu_{Eat}(i_{Act}) = 1$, $\mu_{Sport}(i_{Act}) = 0$ and $\mu_{AlcoholConsumption}(i_{Act}) = 0$. His or her current blood sugar level is considered as very low and very improbably medium: $\mu_{Low}(i_{Cbs}) = 0.91$, $\mu_{Medium}(i_{Cbs}) = 0.09$ and $\mu_{High}(i_{Cbs}) = 0$. His or her last hypoglycemia was five hours ago: $\mu_{Recent}(i_{Lh}) = 0.8$ and $\mu_{Old}(i_{Lh}) = 0.2$. He or she had a short night's sleep: $\mu_{Bad}(i_{Sq}) = 0.6$ and $\mu_{Good}(i_{Sq}) = 0.4$. He or she did not drink enough water during the day: $\mu_{Insufficient}(i_{Wi}) = 0.6$ and $\mu_{Sufficient}(i_{Wi}) = 0.4$. We obtain $\alpha^*_{(Fbs,Low)} = 0.09$, $\alpha^*_{(Fbs,Medium)} = 0.91$ and $\alpha^*_{(Fbs,High)} = 0.6$.

We continue by studying the explainability of a Mamdani fuzzy inference system. It follows from (proposition 5.1) that the semantic justification of a conclusion "Z is O_j^*" by a relevant subset of rule premises leads to the justification of the degree $\alpha^*_{(Z,O_j)}$ by the same objects. We remind that the semantics *possible* of a fuzzy logic expression (conjunction of propositions) e is associated to the joint possibility distribution (relation 1.11) representing e. For the explainability of the inference results of our fuzzy rule-based system, we introduce a threshold $\eta > 0$, which is set according to what is modeled by the rule-base, for the following purpose:

Definition 5.1 *A fuzzy logic expression e is relevantly possible if we have $\alpha_e \geq \eta$. Otherwise, if $\alpha_e < \eta$, e is said to be* not *relevantly possible.*

Given a conclusion (Z, O_j), and the fuzzy degree $\alpha^*_{(Z,O_j)}$ (relation 5.5), we remind that the inferred fuzzy set O^*_j is the truncation of the fuzzy set O_j at the level $\alpha^*_{(Z,O_j)}$. Thus, by justifying $\alpha^*_{(Z,O_j)}$, we justify the inferred result. For this purpose, we select the rule premises that justify the fuzzy degree $\alpha^*_{(Z,O_j)}$ of a conclusion (Z, O_j) according to η:

- If $\alpha^*_{(Z,O_j)} \geq \eta$, we select the rule premises whose conclusion is (Z, O_j) and fuzzy degree is equal to $\alpha^*_{(Z,O_j)}$. Therefore, these premises are relevantly possible.
- If $\alpha^*_{(Z,O_j)} < \eta$, we select the premises of the rules in the set $B_{(Z,O_j)}$. In this case, all these premises are not relevantly possible.

The rule premises justifying the degree $\alpha^*_{(Z,O_j)}$ of a conclusion (Z, O_j) are captured with their semantics, using Notation 2. Formally, the selection is performed by the following formula:

$$\mathrm{J}(Z,O_j) = \begin{cases} \{(p_i, \mathsf{P}, \alpha_{p_i}) \mid R^i = (p_i, c_i) \in B_{(Z,O_j)} \text{ s.t. } \alpha_{R^i} = \alpha^*_{(Z,O_j)}\} & \text{if } \alpha^*_{(Z,O_j)} \geq \eta \\ \{(p_i, \mathsf{P}, \alpha_{p_i}) \mid R^i = (p_i, c_i) \in B_{(Z,O_j)}\} & \text{if } \alpha^*_{(Z,O_j)} < \eta. \end{cases} \tag{5.6}$$

We reminded in Chapter 1 that possibility rules are combined disjunctively. The function J selects the fuzzy rules that achieve the combination, taking into account the threshold. If we use implicative fuzzy rules, which are combined conjunctively, the function J could be adapted to select the fuzzy rules whose conclusion is (Z, O_j) and activation degree is equal to the lowest one.

Example 5.2 *In our example 5.1, we arbitrarily set the threshold η at* 0.1*. For each conclusion, we select the rule premises justifying its associated fuzzy degree using the relation (5.6):*

- $J(Fbs, Low) = \{(p_4, P, 0), (p_5, P, 0), (p_7, P, 0), (p_8, P, 0), (p_{10}, P, 0.09)\}$,
- $J(Fbs, Medium) = \{(p_1, P, 0.91)\}$,
- $J(Fbs, High) = \{(p_{12}, P, 0.6), (p_{13}, P, 0.6)\}$.

The premises of the rules r_2, r_3, r_6, r_9 and r_{11} are not used to justify an inference result.

5.3 Premise reduction functions

In this section, we define two premise reduction functions $\mathscr{R}$ and $\mathscr{C}$ that reduce a premise $p = p_1 \wedge p_2 \wedge \cdots \wedge p_k$ of a possibility rule $R = (p, c)$. Such functions rely on the threshold η.
Given the premise p and using η, we can form two sets of propositions:

- a set of propositions that are relevantly possible:

$$A_p = \Big\{p_j \mid \alpha_{p_j} \geq \eta \text{ for } j = 1, \dots, k\Big\}. \tag{5.7}$$

- a set of propositions that are *not* considered as relevantly possible:

$$B_p = \Big\{p_j \mid \alpha_{p_j} < \eta \text{ for } j = 1, \dots, k\Big\}. \tag{5.8}$$

When p is relevantly possible i.e., $\alpha_p \geq \eta$, the set B_p is empty. When p is *not* relevantly possible, each proposition p_j is either in A_p or in B_p. In this case, A_p may be empty.
In what follows, we define the functions $\mathscr{R}$ and $\mathscr{C}$. Such functions will be applied to the rule premises in the triplets of $J(Z, O_j)$, see equation (5.6), in the next section.

5.3.1 Reduction function $\mathscr{R}$

Given the premise p, the function $\mathscr{R}$ returns the structure responsible for α_p, which is the conjunction of propositions that make p relevantly possible or not:

$$\mathscr{R}(p) = \begin{cases} p & \text{if } \alpha_p \geq \eta \\ \bigwedge_{p_j \in B_p} p_j & \text{if } \alpha_p < \eta \end{cases}. \tag{5.9}$$

With respect to the threshold η, if p is relevantly possible, $\mathscr{R}$ returns p as it is. Otherwise, if $\alpha_p < \eta$, $\mathscr{R}$ reduces p to the conjunction of the propositions in the set B_p, see (5.8).
Note that $\alpha_{\mathscr{R}(p)} = \alpha_p$: by applying $\mathscr{R}$ to p, the fuzzy degree of p is preserved.

If $\alpha_p < \eta$, we see that p may be reduced in another way: we can extract its propositions that are relevantly possible (if any). This is done by the second premise reduction function $\mathscr{C}$ that we introduce.

Example 5.3 *In our example 5.1, let us apply $\mathscr{R}$ to p_1, p_4, p_5, p_7, p_8, p_{10}, p_{12} and p_{13}: $\mathscr{R}(p_1) = p_1$, $\mathscr{R}(p_4) = (Act, Sport)$, $\mathscr{R}(p_5) = p_5$, $\mathscr{R}(p_7) = (Act, AlcoholConsumption)$, $\mathscr{R}(p_8) = p_8$, $\mathscr{R}(p_{10}) = (Cbs, Medium)$, $\mathscr{R}(p_{12}) = p_{12}$, and $\mathscr{R}(p_{13}) = p_{13}$.*

5.3.2 Reduction function $\mathscr{C}$

Intuitively, with respect to the threshold η and for a premise p that is not relevantly possible, $\mathscr{C}$ returns a conjunction of propositions, called an *unexpectedness*, which is not involved in the determination of α_p, although relevantly possible.

If $\alpha_p < \eta$ and there exist relevantly possible propositions of p, i.e. $A_p \neq \emptyset$, the function $\mathscr{C}$ returns the conjunction of such propositions:

$$\mathscr{C}(p) = \bigwedge_{p_j \in A_p} p_j. \tag{5.10}$$

If $\alpha_p < \eta$, each proposition p_j composing p appears either in $\mathscr{R}(p)$ or in $\mathscr{C}(p)$, according to its fuzzy degree α_{p_j}.

Example 5.4 *In our example 5.1, we apply* $\mathscr{C}$ *to* p_4 *and* p_{10}*:* $\mathscr{C}(p_4) = \mathscr{C}(p_7) = (Cbs, Low)$ *and* $\mathscr{C}(p_{10}) = (Lh, Recent)$*.*

5.4 Justification and unexpectedness of a conclusion (Z, O_j)

For a conclusion (Z, O_j), by applying the premise reduction function $\mathscr{R}$ to the premises of the triplets in $J(Z, O_j)$, see (5.6), we obtain a justification of the conclusion (Z, O_j):

$$\text{Justification}(Z, O_j) = \Big\{(\mathscr{R}(p), sem, d) \mid (p, sem, d) \in J(Z, O_j)\Big\}. \tag{5.11}$$

The set Justification(Z, O_j) contains fuzzy logic expressions which are sufficient to justify "Z is O_j^* is $\alpha^*_{(Z,O_j)}$-possible".

Example 5.5 *For each conclusion of our Mamdani system (Example 5.1), we form its justification:*

- *Justification*$(Fbs, Low) = \{(\mathscr{R}(p_4), \mathsf{P}, 0), (\mathscr{R}(p_5), \mathsf{P}, 0), (\mathscr{R}(p_7), \mathsf{P}, 0), (\mathscr{R}(p_8), \mathsf{P}, 0), (\mathscr{R}(p_{10}), \mathsf{P}, 0.09)\}$*, which could be interpreted as:*

  > *"The blood sugar level will not be low because the activity is not sport or alcohol consumption, and the current blood sugar level is not medium."*

- *Justification*$(Fbs, Medium) = \{(\mathscr{R}(p_1), \mathsf{P}, 0.91)\}$*. It could be in natural language:*

"It is very possible that the blood glucose level will be medium, as the chosen activity is eating and the current blood glucose level is considered low."

- *Justification*$(Fbs, High) = \{(\mathscr{R}(p_{12}), \mathsf{P}, 0.6), (\mathscr{R}(p_{13}), \mathsf{P}, 0.6)\}$. *From this extraction, a natural language explanation could be:*

  "It is possible that the blood sugar level will be high because he did not drink enough water and did not get enough sleep last night."

If $\alpha^*_{(Z,O_j)} < \eta$, we can obtain unexpectedness of (Z, O_j) by applying, when it is possible, the reduction function $\mathscr{C}$ to the premise of each triplet in $J(Z, O_j)$:

$$\text{Unexpectedness}(Z, O_j) = \Big\{(\mathscr{C}(p), sem, \alpha_{\mathscr{C}(p)}) \mid (p, sem, d) \in J(Z, O_j), d < \eta, A_p \neq \emptyset\Big\}. \quad (5.12)$$

The fuzzy logic expressions in Unexpectedness(Z, O_j) are not involved in the determination of $\alpha^*_{(Z,O_j)}$. Given an unexpectedness $\{(p', sem, d)\}$, we can form statements such as "even if p' is relevantly possible, Z is O^*_j is not possible".

Example 5.6 *In our example 5.1, we extract the unexpectedness of the conclusion* (Fbs, Low). *We get:*

$$Unexpectedness(Fbs, Low) = \{(\mathscr{C}(p_4), \mathsf{P}, 0.91), (\mathscr{C}(p_{10}), \mathsf{P}, 0.8)\}.$$

This could be interpreted in natural language as:

"The blood glucose level will not be low even if the last hypoglycemia occurred very recently and the current blood glucose level is considered very low."

5.5 Conclusion

In this chapter, we have studied how to justify the inferred conclusions of a Mamdani fuzzy inference system. The method for generating explanations of fuzzy inference decisions is similar to the one proposed for possibilistic rule-based systems (Chapter 4).
We first investigated the semantics of the total inferred conclusion of a Mamdani system. For a conclusion (Z, O_j) of the fuzzy system, we then justified the highest activation degree among all activation degrees of the fuzzy rules of the conclusion (Z, O_j) by an adequate selection of rule premises. Therefore, we justified the inferred fuzzy set O^*_j. We then applied reduction functions to these selected premises, in order to form two explanations: the justification of a conclusion and its unexpectedness. Such extractions rely on the threshold η, which states if a fuzzy logic expression is relevantly possible. Therefore the

change of this parameter leads to variations on the content of the explanations. It must be set according to what the rule base models.

For our explanation generation processing chain, the methods introduced in this chapter allow us to extract the content of explanations of fuzzy inference decisions, which will be represented in terms of conceptual graphs in Part D. The semantics associated to each fuzzy logic expression must be captured, in order to represent graphically the semantics associated to each represented fuzzy value.

PART D

Representation of explanations of possibilistic and fuzzy rule-based system inference decisions

For our processing chain for producing natural language explanations (Figure 1), we propose to represent by conceptual graphs, explanations of the inference results of two rule-based systems: a possibilistic rule-based system and a Mamdani fuzzy inference system (a fuzzy rule-based system composed of possibility rules). We rely on the explanation extraction methods developed for these two systems that we introduced in Chapters 4 and 5, respectively.
In the following, we begin by introducing a general method for representing explanations (Chapter 6). Then, we extend it to represent explanations of possibilistic inference decisions (Chapter 7) and explanations of fuzzy inference decisions (Chapter 8). For each type of rule-based system, we represent three explanations of an inference result: its justification, its unexpectedness and a combination of its justification and its unexpectedness. Finally, our constructions are illustrated by explanations of the inference results of the rule-based systems used in the examples in Chapters 4 and 5.

Chapter 6

A framework for the representation of explanations

In this chapter, we elaborate a framework for representing explanations in terms of conceptual graphs. We begin by setting out the objects that let us form an explanation of an inference decision of a rule-based system. We state that an explanation is composed of:

- $m + 1$ statements ($m \geq 1$), where one is an *observed phenomenon*. The other m statements are related to this phenomenon, and can be either justifications for the phenomenon, or unexpectedness that do not prevent the phenomenon from occurring.
- a link between the phenomenon and the other m statements to structure the explanation. For example, such link may be denoted "isJustifiedBy" or "evenIf".

The *representation of an explanation* in terms of conceptual graphs is achieved as follows. Given an explanation, each of its statements is represented by a conceptual graph. To structure the explanation, the graphs representing the statements are nested in a root conceptual graph that contains a relation node representing the link between the statements. The resulting *nested conceptual graph* is a representation of the explanation.

In what follows, we start by giving the minimal vocabulary (Section 6.1) to build the root conceptual graph R of the representation. Then, assuming that the minimal vocabulary is extended in order to represent the $m+1$ statements by the conceptual graphs denoted $D, N_1, N_2, \ldots, N_m$, we give the interpretation of each of the graphs composing the representation (Section 6.2). Then, in Section 6.3, we define the nested conceptual graph representing the explanation. Finally, our construction is illustrated with an example of an explanation (Section 6.4) and we conclude (Section 6.5).

In Chapters 7 and 8, this framework is used to represent explanations of possibilistic inference decisions and explanations of fuzzy inference decisions respectively. The content of the represented explanations is extracted using the methods (justification and unexpectedness) developed in Chapters 4 and 5 respectively. From these extractions, we form the statements composing a considered explanation and use the framework presented in this chapter to obtain a representation of this explanation.

6.1 Vocabulary

Let us define a minimal vocabulary $\mathcal{V}^0 = (T_C, T_R, \mathcal{I}, \tau, \sigma)$ to represent the root conceptual graph of the representation. We remind that T_C is the set of concept types, T_R is the set of relation symbols, $\mathcal{I}$ is the set of individual markers, $\tau : \mathcal{I} \to T_C$ is an individual typing function and σ is a relation symbol signature, which gives for each relation symbol of T_R the concept type of each of its arguments [34].
The minimal vocabulary contains the objects for structuring the explanation. It has two concept types: Phenomenon and $\mathtt{e}$, a relation symbol $\mathtt{t}$, and $m + 1$ individual markers that are named Statements. For an explanation, Phenomenon and $\mathtt{e}$ are types of statements and the relation symbol $\mathtt{t}$ is the link between the statement of type Phenomenon and the other m statements of type $\mathtt{e}$. The concept type $\mathtt{e}$ and the relation symbol $\mathtt{t}$ will be set according to the explanation that we represent.
In a conceptual graph based on $\mathcal{V}^0$, we may find a concept node of type Phenomenon and marker Statement_0, m concept nodes of type $\mathtt{e}$ and marker Statement_i for $i = 1, 2, \ldots, m$ and a relation node of type $\mathtt{t}$, which will be linked by multi-edges to the $m + 1$ concept nodes that we just described.

Definition 6.1 *We define $\mathcal{V}^0$ as follows:*

- $T_C = \{\mathtt{e}\} \cup \{\textit{Phenomenon}\}$,
- $T_R = \{\mathtt{t}\}$ *such that arity*$(\mathtt{t}) = m + 1$,
- $\mathcal{I} = \{\textit{Statement}_0, \textit{Statement}_1, \ldots, \textit{Statement}_m\}$ *with card*$(\mathcal{I}) = m + 1$,
- $\tau : \mathcal{I} \to T_C$ *such that:*
  - $\textit{Statement}_0 \mapsto \textit{Phenomenon}$,
  - $\textit{Statement}_i \mapsto \mathtt{e}$ *for* $i = 1, 2, \ldots, m$.
- *The signature map σ is given by:*
  - $\sigma(\mathtt{t}) = (\textit{Phenomenon}, \mathtt{e}, \mathtt{e}, \ldots, \mathtt{e})$.

6.2 Graphs of the representation

We consider a vocabulary $\mathcal{V}^{\mathcal{E}}$ that extends the minimal vocabulary $\mathcal{V}^0$ in order to construct $m + 2$ conceptual graphs D, N_1, N_2, ..., N_m, R. The graphs D, N_1, N_2, ..., N_m represent the $m + 1$ statements with the following interpretations:

Definition 6.2 *D is a conceptual graph built on $\mathcal{V}^{\mathcal{E}}$. It is a graphical representation of a statement, which describes an observed phenomenon.*

Definition 6.3 *For $i = 1, 2, \cdots, m$, each N_i is a conceptual graph built on $\mathcal{V}^{\mathcal{E}}$ that represents graphically a statement related to the phenomenon represented by D. For instance, it can be a statement that justifies the phenomenon or an unexpectedness statement.*

We define the root conceptual graph R of the representation as follows:

Definition 6.4 *The graph R is the star BG (Definition 1.8) built on $\mathcal{V}^{\mathcal{E}}$ where the unique relation node r is of type t and the $m+1$ concept nodes are noted $c_0, c_1, \ldots, c_m$, where c_0 is of type Phenomenon and $c_1, c_2, \ldots, c_m$ are of type e. Their individual markers are respectively Statement$_0$, Statement$_1$, ..., Statement$_m$. The multi-edges are labeled (r, j, c_j) for $j = 0, 1, \ldots, m$. R structures the explanation by representing the link between the observed phenomenon D and the statements $N_1, N_2, \ldots, N_m$.*

We note that R can be built on the minimal vocabulary $\mathcal{V}^0$. An example of a graph R is given in Figure 6.1.

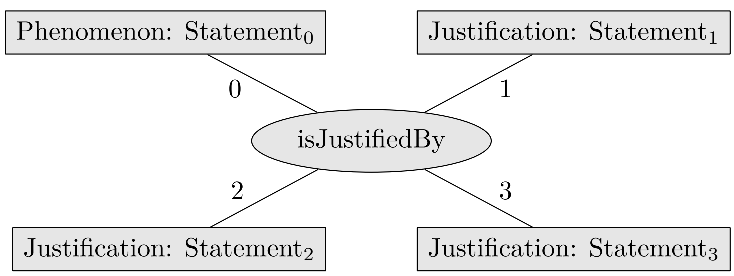


Figure 6.1: The graph R, with $m = 3$ and where e and t are noted “Justification” and “isJustifiedBy” respectively.

We note that R can be adapted to represent more complex explanations. For example, R can be extended in order to represent the combination of two explanations that have the same observed phenomenon D. This will be shown in the next chapters.

6.3 Representation of an explanation

We define the representation of an explanation as a nested conceptual graph G defined by its associated tree as in [34], which is denoted $Tree(G) = (V_T, U_T, l_T)$. For our representation, the conceptual graphs $D, N_1, N_2, \dots, N_m$ are nested in the concept nodes of R:

Definition 6.5 *The tree associated to the NBG G, denoted $Tree(G) = (V_T, U_T, l_T)$ is given by:*
- *$V_T = \{R, D, N_1, N_2, \dots, N_m\}$ is the set of nodes,*
- *$U_T = \{(R, D), (R, N_1), (R, N_2), \dots, (R, N_m)\}$ 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 $i = 1, 2, \dots, m$.*

An example of the representation of an explanation, where e and t are set is given in Figure 6.2.

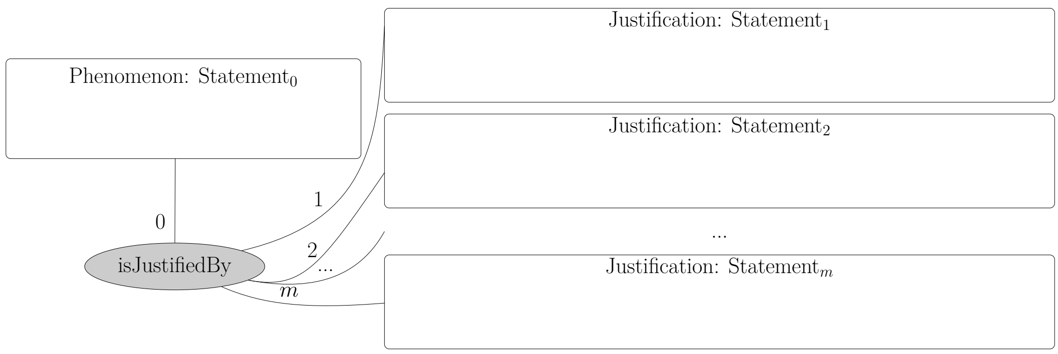


Figure 6.2: Representation of an explanation where e and t are noted "Justification" and "isJustifiedBy" respectively.

6.4 Example

Let us represent the following natural language explanation in terms of conceptual graphs:

Alice owes 20 euros to Bob because Alice borrowed 20 euros from Bob.

In our case, the explanation is formed by two statements ($m = 1$). "Alice owes 20 euros to Bob" is the observed phenomenon, which is justified by the statement "Alice borrowed 20 euros from Bob". To represent this explanation, we introduce an ad hoc vocabulary $V^{\mathcal{E}}$ that extends V^0. We set e and t as

"Justification" and "isJustifiedBy" respectively. We add Alice, Bob and the value 20 to the set of individual markers. Their respective concept types are Human, Human and Amount. Two relation symbols of arity 3 denoted owe and borrow are introduced. We establish the vocabulary $V^{\mathcal{E}}$ as follows:

- $T_C = \{\text{Justification}, \text{Phenomenon}, \text{Human}, \text{Amount}\}$,
- $T_R = \{\text{isJustifiedBy}, \text{owe}, \text{borrow}\}$ such that $\text{arity}(\text{isJustifiedBy}) = 2$ and $\text{arity}(\text{owe}) = \text{arity}(\text{borrow}) = 3$,
- $\mathcal{I} = \{\text{Statement}_0, \text{Statement}_1, \text{Alice}, \text{Bob}, 20\}$,
- $\tau : \mathcal{I} \to T_C$ such that:
  - $\text{Statement}_0 \mapsto \text{Phenomenon}$,
  - $\text{Statement}_1 \mapsto \text{Justification}$,
  - $\text{Alice} \mapsto \text{Human}$,
  - $\text{Bob} \mapsto \text{Human}$,
  - $20 \mapsto \text{Amount}$.
- The signature map σ is given by:
  - $\sigma(\text{isJustifiedBy}) = (\text{Phenomenon}, \text{Justification})$,
  - $\sigma(\text{owe}) = \sigma(\text{borrow}) = (\text{Human}, \text{Amount}, \text{Human})$.

We define $D = (C_D, R_D, E_D, l_D)$ and $N_1 = (C_N, R_N, E_N, l_N)$ as star BGs.
The graph D is composed of an unique relation node r_D of type owe and three concept nodes c_1, c_2 and c_3 of type Human, Amount, Human and marker Alice, 20, Bob. Similarly, the graph N_1 is composed of an unique relation node r_N of type borrow and three concept nodes c'_1, c'_2 and c'_3 of type Human, Amount, Human and marker Alice, 20, Bob.
For D (resp. N_1) the multi-edges are labeled $(r_D, 0, c_1), (r_D, 1, c_2)$ and $(r_D, 2, c_3)$ (resp. $(r_N, 0, c'_1), (r_N, 1, c'_2)$ and $(r_N, 2, c'_3)$). We form a root graph R (definition 6.4) and represent the explanation using Definition 6.5. The obtained representation is given in Figure 6.3.

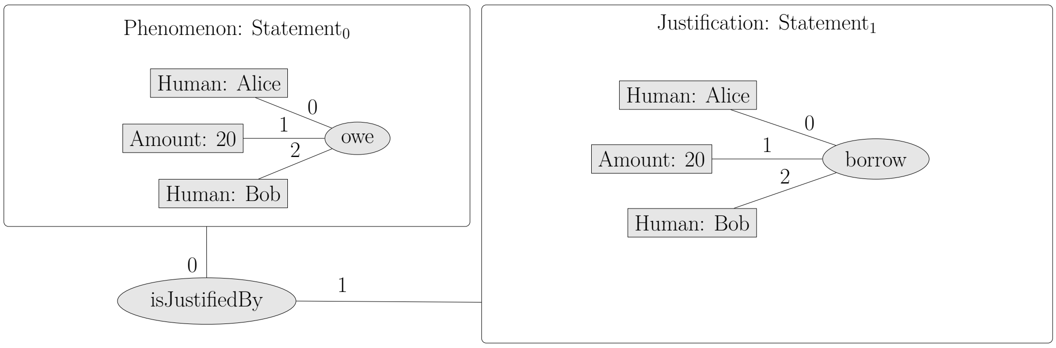


Figure 6.3: Representation of an explanation composed of two statements.

6.5 Conclusion

In this chapter, we elaborated a framework to represent explanations in terms of conceptual graphs. The representation of the explanation is composed of $m+2$ conceptual graphs $D, N_1, N_2, \dots, N_m, R$, where $D, N_1, N_2, \dots, N_m$ represent each of the $m+1$ statements composing the explanation and R represents the structure of the explanation. The resulting representation is a nested conceptual graph, which we have defined by its associated tree. Finally, we have illustrated our construction by representing a simple natural language explanation.

In the following, we rely on this framework to represent explanations of possibilistic inference decisions (chapter 7) and explanations of fuzzy inference decisions (chapter 8). The content of these explanations is extracted using the methods introduced in chapters 4 and 5, respectively.

Chapter 7

Representation of explanations of possibilistic inference decisions

The work in this chapter has led to the publication of a conference paper: Baaj, I., Poli, J. P., Ouerdane, W. & Maudet, N. (2021, September). Representation of Explanations of Possibilistic Inference Decisions. In 2021 European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU).

In this chapter, we represent graphically two explanations: the justification and the unexpectedness of the possibility degree $\pi^*_{b(x)}(u)$ of an output attribute value u (see equations (4.21) and (4.22) from Chapter 4). To represent these explanations, we rely on our framework introduced in Chapter 6. The resulting conceptual graphs are visual representations of the outcomes of several analytical operations performed on the possibilistic rule base that constitute explanations.

Using our framework presented in Chapter 6, we start by specifying, in the context of an explanation of a possibilistic inference decision, the objects that compose it. For both explanations (justification and unexpectedness), the observed phenomenom is the possibility degree $\pi^*_{b(x)}(u)$ of an output attribute value u. The other statements are the possible or certain possibilistic expressions captured in the considered explanation.
This chapter is structured as follows. In Section 7.1, we introduce the notion of a *possibilistic conceptual graph*, which is defined as a conceptual graph where each concept node is gifted with a degree and a semantics. Such a graph lets us represent graphically the observed phenomenon or each of the possible or certain possibilistic expressions captured in a justification or in an unexpectedness.
To build the vocabulary to represent an explanation, we must first perform a

preprocessing step based on a justification or unexpectedness (Section 7.2). In this step, we define a *possibilistic explanation query*, which is a structure that can capture a justification or an unexpectedness and determine the explanation statements. Starting from a possibilistic explanation query, we establish a mapping between the attribute and the attribute subdomain underlying each of the propositions composing the possibilistic expressions of the explanation. Finally, we end this section by defining two explicit possibilistic explanation queries: one for the justification of $\pi^*_{b(x)}(u)$ and the other for its unexpectedness.

Each possibilistic explanation query gives rise to a vocabulary (Section 7.3). From this vocabulary, we construct all the graphs composing the representation of an explanation (Section 7.4). Then, the representation of the explanation is achieved by nesting the possibilistic conceptual graphs representing the statements in the root conceptual graph, according to Definition 6.5 presented in Chapter 6.

In Section 7.5, we extend the framework in order to represent an explanation, which is the combination of the justification and the unexpectedness of $\pi^*_{b(x)}(u)$. To build such a representation, we combine the two possibilistic explanation queries associated respectively with justification and unexpectedness into a new possibilistic explanation query. Then, we define a new root graph for structuring this explanation.

All our constructions are illustrated (Section 7.6) by the explanations extracted from the two possibilistic rule-based systems used in the examples in Chapter 4. Finally, we conclude with some perspectives (Section 7.7).

7.1 Possibilistic conceptual graphs

We introduce possibilistic conceptual graphs that extend basic conceptual graphs (BG) [34] by adding two additional fields to the labels of concept nodes:

Definition 7.1 *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 $c \in C$:*

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

where $sem_c \in \{\mathsf{P}, \mathsf{C}\}$ and $d_c \in [0, 1]$.

The definition of a *star BG* (Definition 1.8) i.e., a BG restricted to a relation node and its neighbors, is naturally extended as a star PCG.

7.2 Possibilistic explanation query

To describe the vocabulary for representing an explanation, we introduce the notion of a *possibilistic explanation query*:

> **Definition 7.2** *A possibilistic 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 a possibilistic explanation query $\mathcal{E} = (\mathcal{T}, b, u)$, where $m = \text{card}(\mathcal{T}) \geq 1$. The possibilistic explanation query $\mathcal{E}$ is the input for representing an explanation of a possibilistic inference decision composed of $m + 1$ statements. The statement that is the observed phenomenom will be constructed using the output attribute b, the value u and the possibility degree $\pi^*_{b(x)}(u)$. We adopt the following notations for $\mathcal{E}$:

> **Notation 3** *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* $p^{(i)} = p^{(i)}_1 \wedge p^{(i)}_2 \wedge \cdots \wedge p^{(i)}_{k_i}$ *where for each* $j = 1, 2, \cdots, k_i$ *we have:*
> • $p^{(i)}_j$ *is the proposition "*$a^{(i)}_j(x) \in P^{(i)}_j$*", where* $a^{(i)}_j$ *is an attribute with a normalized possibility distribution* $\pi_{a^{(i)}_j} : D_{a^{(i)}_j} \to [0, 1]$, $P^{(i)}_j \subseteq D_{a^{(i)}_j}$ *and* x *is an item.*
> • $\mathcal{A}^{(i)} = \{a^{(i)}_1, a^{(i)}_2, \cdots, a^{(i)}_{k_i}\}$ *with card*$(\mathcal{A}^{(i)}) = k_i$,
> • $\mathcal{S}^{(i)} = \{P^{(i)}_1, P^{(i)}_2, \cdots, P^{(i)}_{k_i}\}$ *with card*$(\mathcal{S}^{(i)}) = k_i$.

We take the *disjoint unions*:

$$\mathcal{A} = \dot{\bigcup}_{1 \leq i \leq m} \mathcal{A}^{(i)} \text{ and } \mathcal{S} = \dot{\bigcup}_{1 \leq i \leq m} \mathcal{S}^{(i)}.$$

These disjoint unions will allow us to define an application τ: $\mathcal{S} \to \mathcal{A}$ verifying $\tau(P^{(i)}_j) = a^{(i)}_j$ and are necessary because the domains of two distinct attributes $a^{(i)}_j$ and $a^{(i')}_{j'}$ with $i \neq i'$ may have a non-empty intersection. Therefore, the sets $P^{(i)}_j$ and $P^{(i')}_{j'}$ of the two propositions $p^{(i)}_j$ and $p^{(i')}_{j'}$ may be equal.
From the content extracted for justifying the possibility degree $\pi^*_{b(x)}(u)$ of an output attribute value u, see (4.21) and its unexpectedness, see (4.22), we take the following explanation queries:

$$\mathcal{E}_J = (\text{Justification}_{b(x)}(u), b, u) \tag{7.1a}$$

and

$$\mathcal{E}_U = (\text{Unexpectedness}_{b(x)}(u), b, u). \tag{7.1b}$$

From each, we will be able to construct an associated vocabulary.

7.3 Vocabulary construction

Let $\mathcal{V}^{\mathcal{E}} = (T_C, T_R, \mathcal{I}, \tau, \sigma)$ be the vocabulary associated to the explanation query $\mathcal{E} = (\mathcal{T}, b, u)$ that extends the minimal vocabulary of the framework (Definition 6.1). 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 $p_1^{(i)}, p_2^{(i)}, \cdots, p_{k_i}^{(i)}$.

The vocabulary $\mathcal{V}^{\mathcal{E}}$ includes the objects used to structure the explanation, as in the minimal vocabulary of the framework (Definition 6.1). We explicitly define $\mathcal{V}^{\mathcal{E}}$ as follows:

Definition 7.3 *The vocabulary $\mathcal{V}^{\mathcal{E}}$ is defined by:*
• $T_C = \{b\} \cup \mathcal{A} \cup \{\mathtt{e}\} \cup \{Phenomenon\}$ *with* $card(T_C) = 3 + \sum_{i=1}^{m} k_i$.
• $T_R = \{inferred_{v^{(i)}} | v^{(i)} \in \mathcal{T}\} \cup \{\mathtt{t}\}$ *with* $card(T_R) = m + 1$ *and such that* $arity(inferred_{v^{(i)}}) = k_i + 1$ *and* $arity(\mathtt{t}) = m + 1$.
• $\mathcal{I} = \{\{u\}\} \cup \mathcal{S} \cup \{Statement_0, Statement_1, \cdots, Statement_m\}$ *with* $card(\mathcal{I}) = m + 2 + \sum_{i=1}^{m} k_i$.
• $\tau : \mathcal{I} \rightarrow T_C$ *such that:*

- $\{u\} \mapsto b$,
- $P_j^{(i)} \mapsto a_j^{(i)}$,
- $Statement_0 \mapsto Phenomenon$,
- $Statement_i \mapsto \mathtt{e}$ *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* $v^{(i)} \in \mathcal{T}$,
  - $\sigma(\mathsf{t}) = (Phenomenon, \mathsf{e}, \mathsf{e}, \cdots, \mathsf{e})$.

In the vocabulary $\mathcal{V}^{\mathcal{E}_J}$ associated to the explanation query $\mathcal{E}_J$, see (7.1), the concept type e is noted "Justification" and the relation symbol t is noted "isJustifiedBy". In the vocabulary $\mathcal{V}^{\mathcal{E}_U}$ associated to $\mathcal{E}_U$, see (7.1), we respectively note them "Unexpectedness" and "evenIf".

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

Given a possibilistic explanation query $\mathcal{E} = (\mathcal{T}, b, u)$ (Definition 7.2), 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 type$(c) = b$ and marker$(c) = \{u\}$, we put:

$$sem_c = \mathsf{P} \text{ and } d_c = \pi^*_{b(x)}(u). \tag{7.2}$$

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

$$sem_c = sem^{(i)} \text{ and } d_c = \begin{cases} \pi(p_j^{(i)}) \text{ if } sem^{(i)} = \mathsf{P} \\ n(p_j^{(i)}) \text{ if } sem^{(i)} = \mathsf{C} \end{cases}. \tag{7.3}$$

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 $D, N_1, N_2, \cdots, N_m$:

Definition 7.4 *D is defined as the PCG reduced to a single concept node with label* $(b : \{u\} \mid \mathsf{P}, \pi^*_{b(x)}(u))$.

Definition 7.5 *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 (7.2), (7.3). The multi-edges are labeled $(r_i, 0, c_b^{(i)})$ and $(r_i, j, c_{a_j}^{(i)})$ for $j = 1, 2, \cdots, k_i$.*

In Figure 7.1, we give an example of N_i.

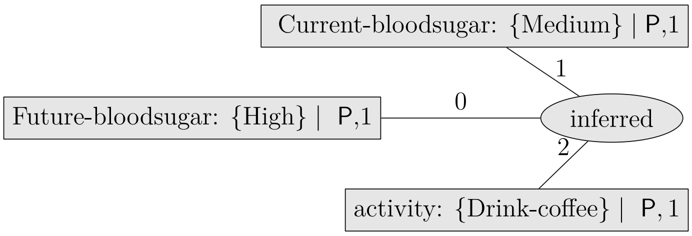


Figure 7.1: Example of a graph N_i.

The root graph R is constructed according to Definition 6.4 of the framework, using the vocabulary $\mathcal{V}^{\mathcal{E}}$.
Finally we use Definition 6.5 to obtain the representation of the explanation of a possibilistic inference decision.

From $\mathcal{V}^{\mathcal{E}_J}$ and $\mathcal{V}^{\mathcal{E}_U}$, we can construct the graphs of the representation of the justification and the unexpected of $\pi^*_{b(x)}(u)$. For each explanation, the nesting of the graphs representing its statements in their respective root graph allows us to obtain the representation of an explanation.

7.5 Representation of the combination of the justification and the unexpectedness

For an output attribute value u and the two possibilistic explanation queries $\mathcal{E}_J = (\mathcal{T}_J, b, u)$ and $\mathcal{E}_U = (\mathcal{T}_U, b, u)$ associated to its justification and its unexpectedness respectively, see (7.1), where $m_J = card(\mathcal{T}_J)$ and $m_U = card(\mathcal{T}_U)$, we can represent an explanation of $m + 1 = m_J + m_U + 1$ statements, which is the combination of the two explanations.
This combination will allow us to formulate new natural language explanations. Given an unexpected $\{(p_1, sem_1, d_1)\}$ of an output attribute value u and its justification $\{(p_2, sem_2, d_2)\}$, we can express an explanation such as: "Even if p_1, $b(x)$ is u is possible to a degree $\pi^*_{b(x)}(u)$ because of p_2".
To construct the representation of the combination, let us re-index the triplets in $\mathcal{T}_J$ from 1 to m_J, i.e., $\mathcal{T}_J = \{v^{(1)}, v^{(2)}, \ldots, v^{(m_J)}\}$ and those of $\mathcal{T}_U$ from $m_J + 1$ to $m = m_J + m_U$ i.e., $\mathcal{T}_U = \{v^{(m_J+1)}, v^{(m_J+2)}, \ldots, v^{(m)}\}$. We form a new possibilistic explanation query $\mathcal{E}_{JU}$ such that:

$$\mathcal{E}_{JU} = (\mathcal{T}_J \cup \mathcal{T}_U, b, u) = (\mathcal{T}, b, u), \tag{7.4}$$

where $\mathcal{T} = \mathcal{T}_J \cup \mathcal{T}_U$ is a disjoint union. From $\mathcal{E}_{JU}$, we adopt Notation 3 and obtain $\mathcal{A}$ and $\mathcal{S}$. We form a new vocabulary, which extends the one presented previously (Definition 7.3) in the following sense: we substitute the concept type e (resp. relation symbol t) by two concept types (resp. relation symbols) named "Justification" and "Unexpectedness" (resp. "isJustifiedBy" of arity $m_J + 1$ and "evenIf" of arity $m_U + 1$). The individual typing function is updated to link statements to their respective types. The signature map is given for the two new relations symbols. We define $\mathcal{V}^{\mathcal{E}_{JU}}$ as follows:

Definition 7.6 *The vocabulary $\mathcal{V}^{\mathcal{E}_{JU}}$ is defined by:*
• $T_C = \{b\} \cup \mathcal{A} \cup \{Phenomenon, Justification, Unexpectedness\}$ *with* $card(T_C) = 4 + \sum_{i=1}^{m} k_i$.
• $T_R = \{inferred_{v^{(i)}} | v^{(i)} \in \mathcal{T}\} \cup \{isJustifiedBy, evenIf\}$ *with* $card(T_R) = m+2$ *and such that* $arity(inferred_{v^{(i)}}) = k_i+1$, $arity(isJustifiedBy) = m_J+1$ *and* $arity(evenIf) = m_U + 1$.
• $\mathcal{I} = \{\{u\}\} \cup \mathcal{S} \cup \{Statement_0, Statement_1, \ldots, Statement_m\}$ *with* $card(\mathcal{I}) = m + 2 + \sum_{i=1}^{m} k_i$.
• $\tau : \mathcal{I} \to T_C$ *such that:*

- $\{u\} \mapsto b$,
- $P_j^{(i)} \mapsto a_j^{(i)}$,
- $Statement_0 \mapsto Phenomenon$,
- $Statement_i \mapsto Justification$ *for* $i = 1, 2, \ldots, m_J$.
- $Statement_i \mapsto Unexpectedness$ *for* $i = m_J + 1, m_J + 2, \ldots, m = m_J + m_U$.

• *The signature map* σ *is given by:*

- $\sigma(inferred_{v^{(i)}}) = (b, a_1^{(i)}, a_2^{(i)}, \cdots, a_{k_i}^{(i)})$ *for* $v^{(i)} \in \mathcal{T}$,
- $\sigma(isJustifiedBy) = (Phenomenon, Justification, \ldots, Justification)$.
- $\sigma(evenIf) = (Phenomenon, Unexpectedness, \ldots, Unexpectedness)$.

We define the graphs $D, N_1, N_2, \ldots, N_m$ using the vocabulary $\mathcal{V}^{\mathcal{E}_{JU}}$ according to Definitions 7.4 and 7.5. We give a new definition of the graph R, which is now a BG that has two relation nodes:

Definition 7.7 *The graph R is composed of two relation nodes r_1 of type isJustifiedBy and r_2 of type evenIf. It has $m + 1$ concept nodes, which are noted $c_0, c_1, \ldots, c_m$, where c_0 is of type "Phenomenon", $c_1, c_2, \ldots, c_{m_J}$ are of type Justification and $c_{m_J+1}, c_{m_J+2}, \ldots, c_m$ are of type Unexpectedness. Their individual markers are respectively $Statement_0$, $Statement_1$, …, $Statement_m$. The multi-edges from the relation node r_1 (respectively r_2) to* *its neighbors are labelled* (r_1, j, c_j) *with* $j = 0, 1, \ldots, m_J$ *(resp.* $(r_2, 0, c_0)$, (r_2, j, c_{m_J+j}) *with* $j = 1, \ldots, m_U$*).*

Finally, we use Definition 6.5 to obtain the representation of this explanation.

7.6 Examples

We illustrate our constructions with the two possibilistic rule based systems used in the examples in Chapter 4.

7.6.1 First example: representations of the explanations of the inference results of the blood sugar control system

We start with Example 4.1, which is a possibilistic rule-based system that controls the blood sugar level of a patient with type 1 diabetes. In this system, we remind that the output values of the output attribute *future-blood-sugar* (fbs) are: *low*, *medium* and *high*. To represent the justification of $\pi^*_{fbs(x)}(low)$, we first form a possibilistic explanation query $\mathcal{E}_J = (\text{Justification}_{fbs(x)}(low), fbs, low)$ according to Definition 7.1. This allows us to elaborate the vocabulary for constructing the graphs of the representation (Definition 7.3):

•T_C = {Future-blood-sugar, Current-blood-sugar, Activity, Justification, Phenomenon}.

• $T_R = \{\text{inferred}, \text{isJustifiedBy}\}$ such that

$$\text{arity(inferred)} = 3 \quad \text{and} \quad \text{arity(isJustifiedBy)} = 2.$$

• $\mathcal{I}$ ={{low},{drunk,drink-coffee,lunch}, {medium,high}, Statement_0, Statement_1}.

• $\tau : \mathcal{I} \to T_C$ such that:

- {low} $\mapsto$ Future-blood-sugar,
- {drunk,drink-coffee,lunch} $\mapsto$ Activity,
- {medium,high} $\mapsto$ Current-blood-sugar,
- $\text{Statement}_0 \mapsto$ Phenomenon,
- $\text{Statement}_1 \mapsto$ Justification.

• The signature map σ is given by:

- $\sigma(\text{inferred}) = (\text{Future-blood-sugar}, \text{Activity}, \text{Current-blood-sugar})$
- $\sigma(\text{isJustifiedBy}) = (\text{Phenomenon}, \text{Justification})$.

We construct the graphs D, N_1 and R according to Definitions 7.4, 7.5 and 6.4 respectively. By nesting D and N_1 in R (Definition 6.5), we obtain the representation of the justification (Figure 7.2).

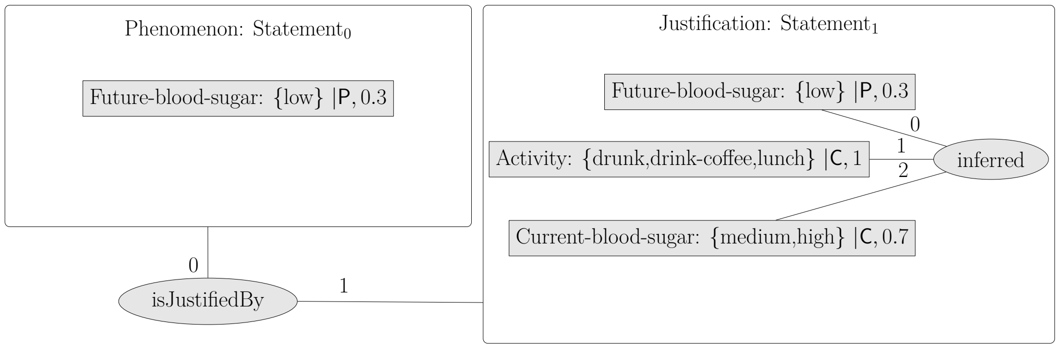


Figure 7.2: Representation of the justification of $\pi^*_{fbs(x)}(low)$.

Similarly, we obtain the representation of justification of the output value *medium* (Figure 7.3). We remind that the extracted justification for *medium* is the same as for *low*.

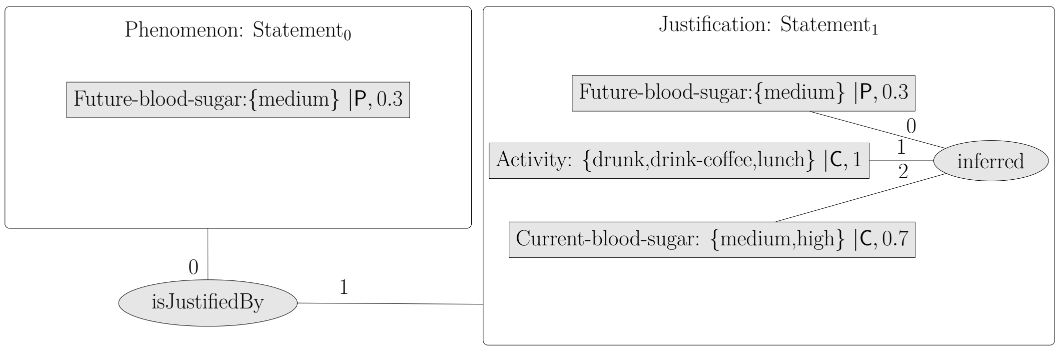


Figure 7.3: Representation of the justification of $\pi^*_{fbs(x)}(medium)$.

We represent the justification of the output attribute value *high* using five graphs (Figure 7.4).

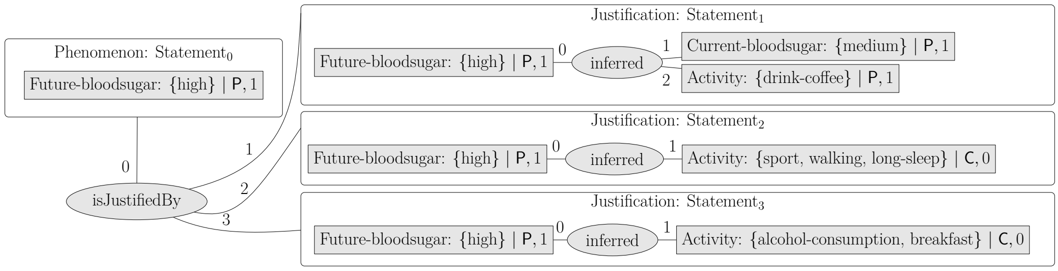


Figure 7.4: Representation of the justification of $\pi^*_{fbs(x)}(high)$.

To represent the unexpectedness of $\pi^*_{fbs(x)}(high)$, we start by establishing its corresponding possibilistic explanation query $\mathcal{E}_U = (\text{Unexpectedness}_{fbs(x)}(high)$, fbs, $high)$ according to Definition 7.1. This allows us to elaborate a vocabulary (Definition 7.3), which is as follows:

- T_C = {Future-blood-sugar, Current-blood-sugar, Unexpectedness, Phenomenon}.
- $T_R = \{\text{inferred}, \text{evenIf}\}$ such that arity(inferred) = 2 and arity(evenIf) = 2.
- $\mathcal{I} = \{\{\text{high}\}, \{\text{low,medium}\}, \text{Statement}_0, \text{Statement}_1\}$.
- $\tau : \mathcal{I} \rightarrow T_C$ such that:
  - {high} $\mapsto$ Future-blood-sugar,
  - {low,medium} $\mapsto$ Current-blood-sugar,
  - Statement$_0$ $\mapsto$ Phenomenon,
  - Statement$_1$ $\mapsto$ Unexpectedness.
- The signature map σ is given by:
  - $\sigma(\text{inferred}) = (\text{Future-blood-sugar}, \text{Current-blood-sugar})$
  - $\sigma(\text{evenIf}) = (\text{Phenomenon}, \text{Unexpectedness})$.

We construct the graphs of the representation using Definitions 7.4, 7.5 and 6.4 and nest the graphs representing the statements in the root graph (Definition 6.5). We obtain the representation of the unexpectedness for the value *high* (Figure 7.5).

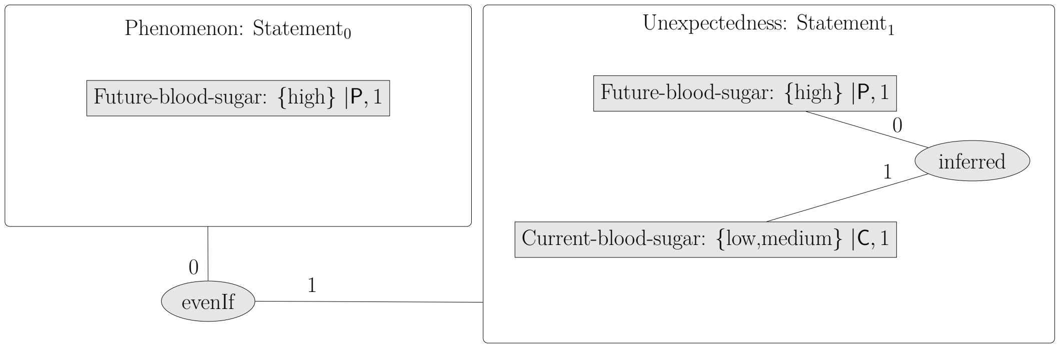


Figure 7.5: Representation of the unexpectedness of $\pi^*_{fbs(x)}(high)$.

To represent the combination of the justification and the unexpectedness of $\pi^*_{fbs(x)}(high)$, we use the equation 7.4 to combine their two associated possibilistic explanation queries into one. From this, we obtain a new vocabulary (Definition 7.6) to represent the explanation. The graph R of this explanation is elaborated with two relation nodes and five concept nodes according to Definition 7.7. We represent R in Figure 7.6.

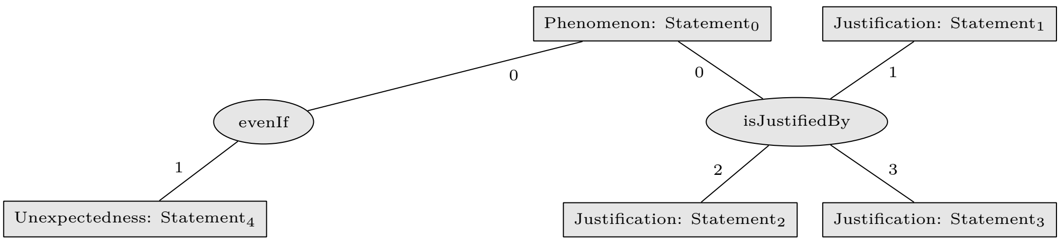


Figure 7.6: Root graph R for the representation of the combination of the justification and the unexpectedness of $\pi^*_{fbs(x)}(high)$.

The combination of the justification of $\pi^*_{fbs(x)}(high)$ and its unexpectedness is represented in Figure 7.7.

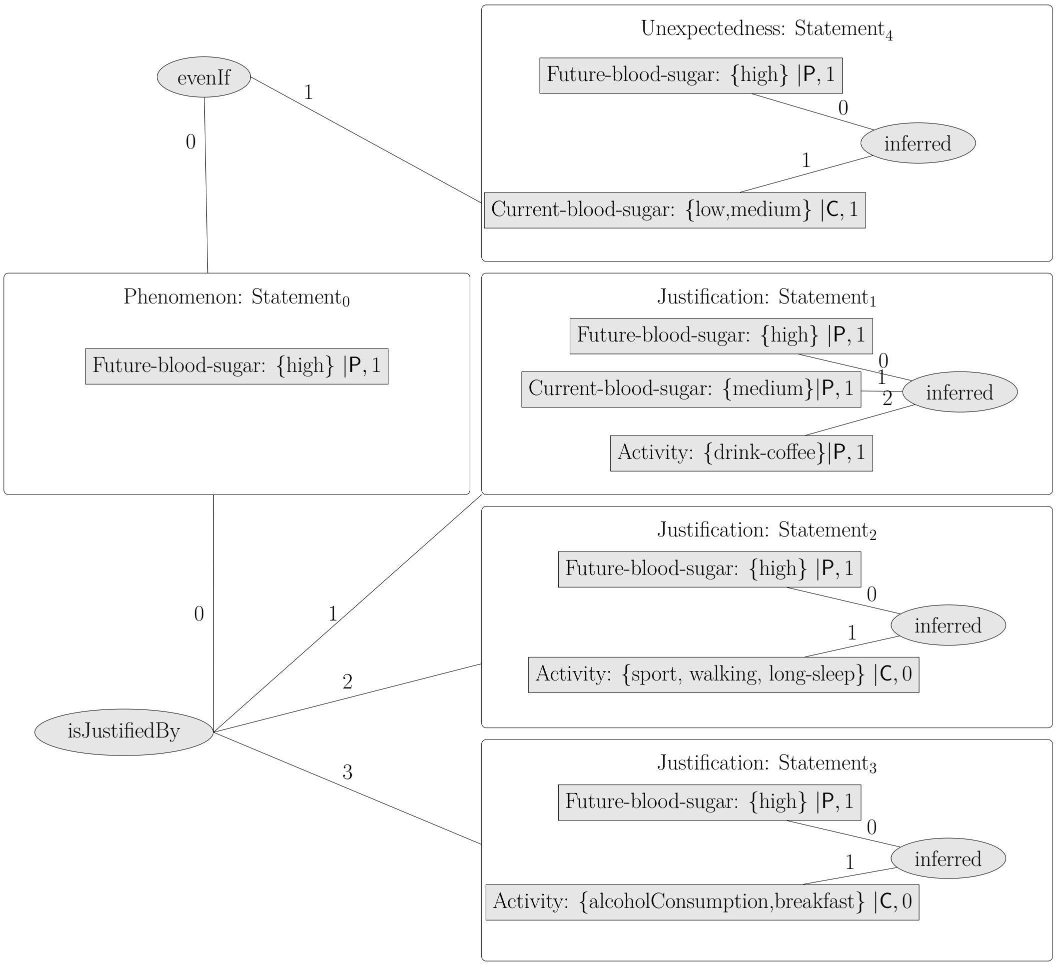


Figure 7.7: Representation of the combination of the justification and the unexpectedness of $\pi^*_{fbs(x)}(high)$.

7.6.2 Second example: representation of the explanations of the inference results of the insulin dose delivery system

In the following, using the example of the possibilistic rule-based system that determine the insulin dose needed by a patient (Section 4.7), we represent graphically explanations of its inference results. The output attribute $insulin-dose$ (id) has three values *low*, *medium* and *high*. The justification of the possibility degree of *low* and *high* are represented in Figures 7.8 and 7.9 respectively. There is no justification for $\pi^*_{id(x)}(medium)$.

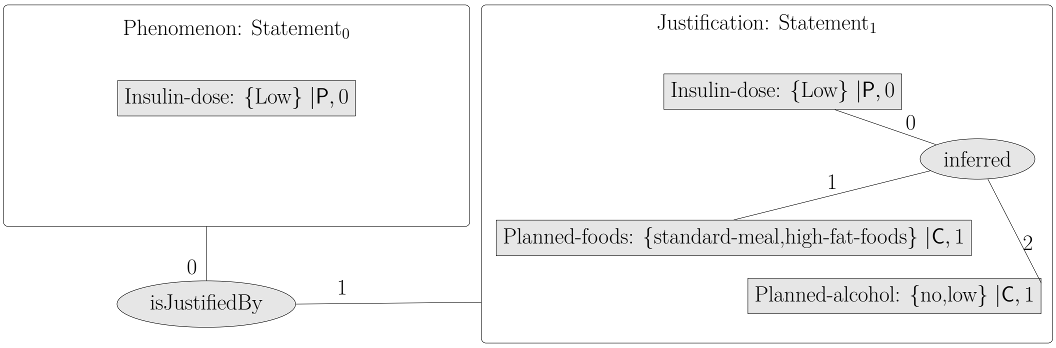


Figure 7.8: Representation of the justification of $\pi^*_{id(x)}(low)$.

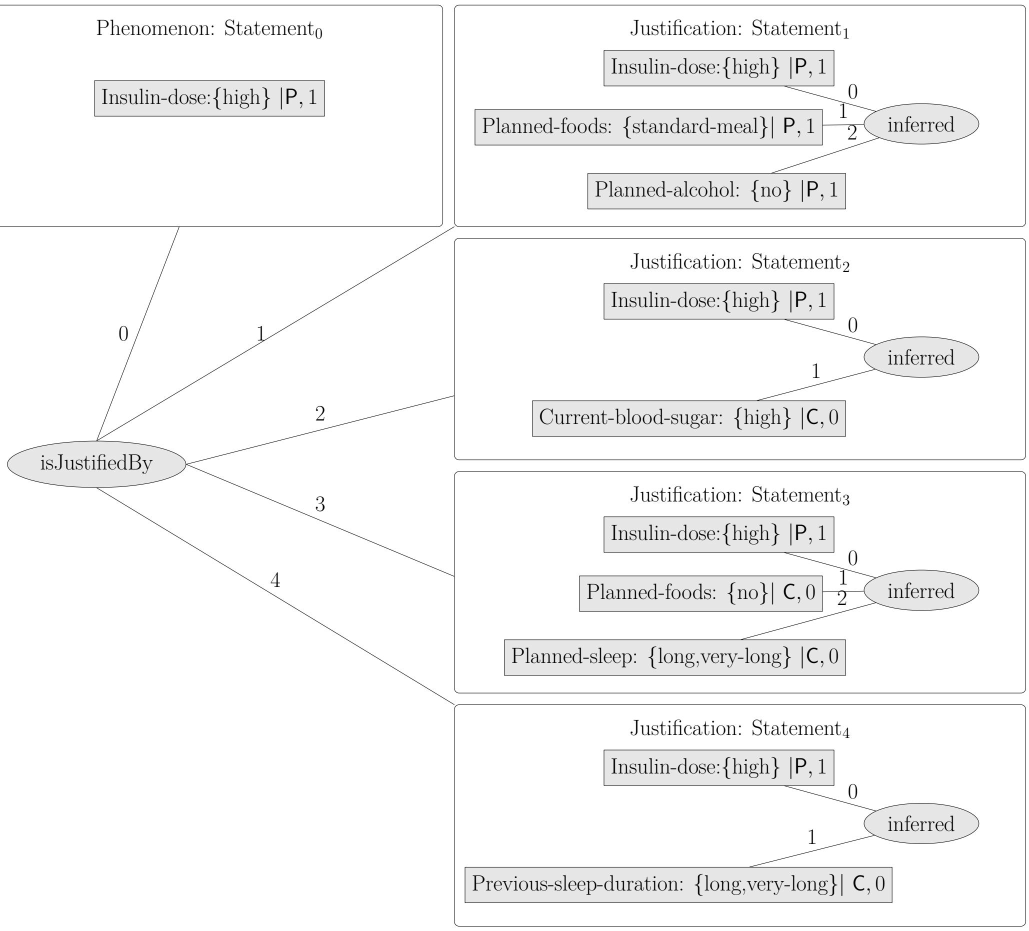


Figure 7.9: Representation of the justification of $\pi^*_{id(x)}(high)$.

The unexpectedness of $\pi^*_{id(x)}(high)$ is represented in Figure 7.10.

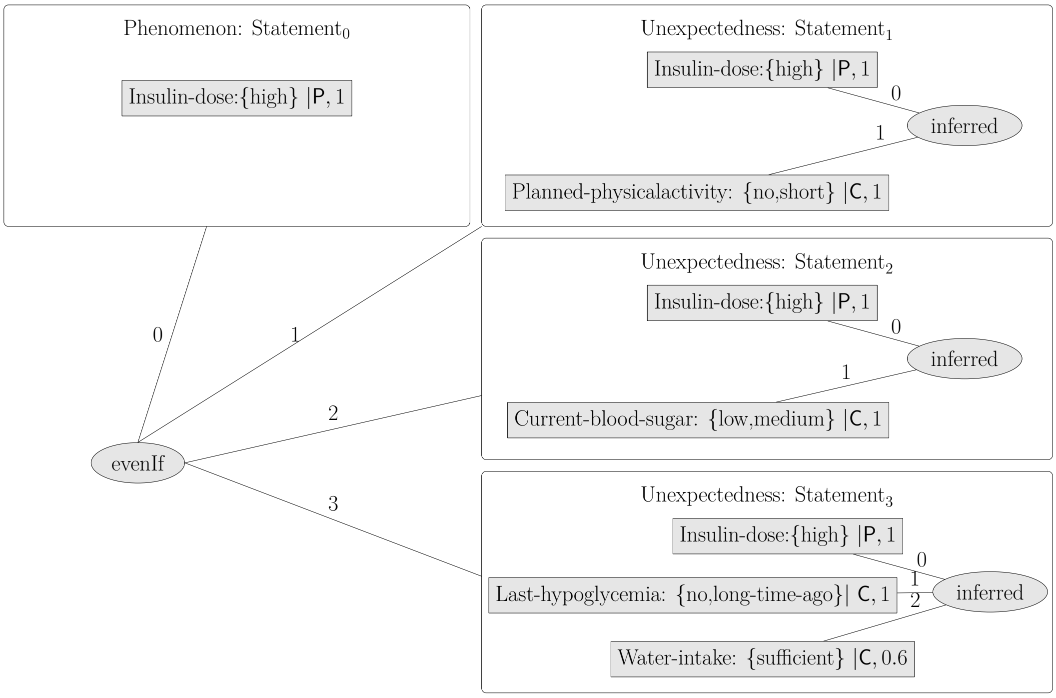


Figure 7.10: Representation of the unexpectedness of $\pi^*_{id(x)}(high)$.

Finally, the combination of the justification and the unexpectedness of $\pi^*_{id(x)}(high)$ is represented in Figure 7.11.

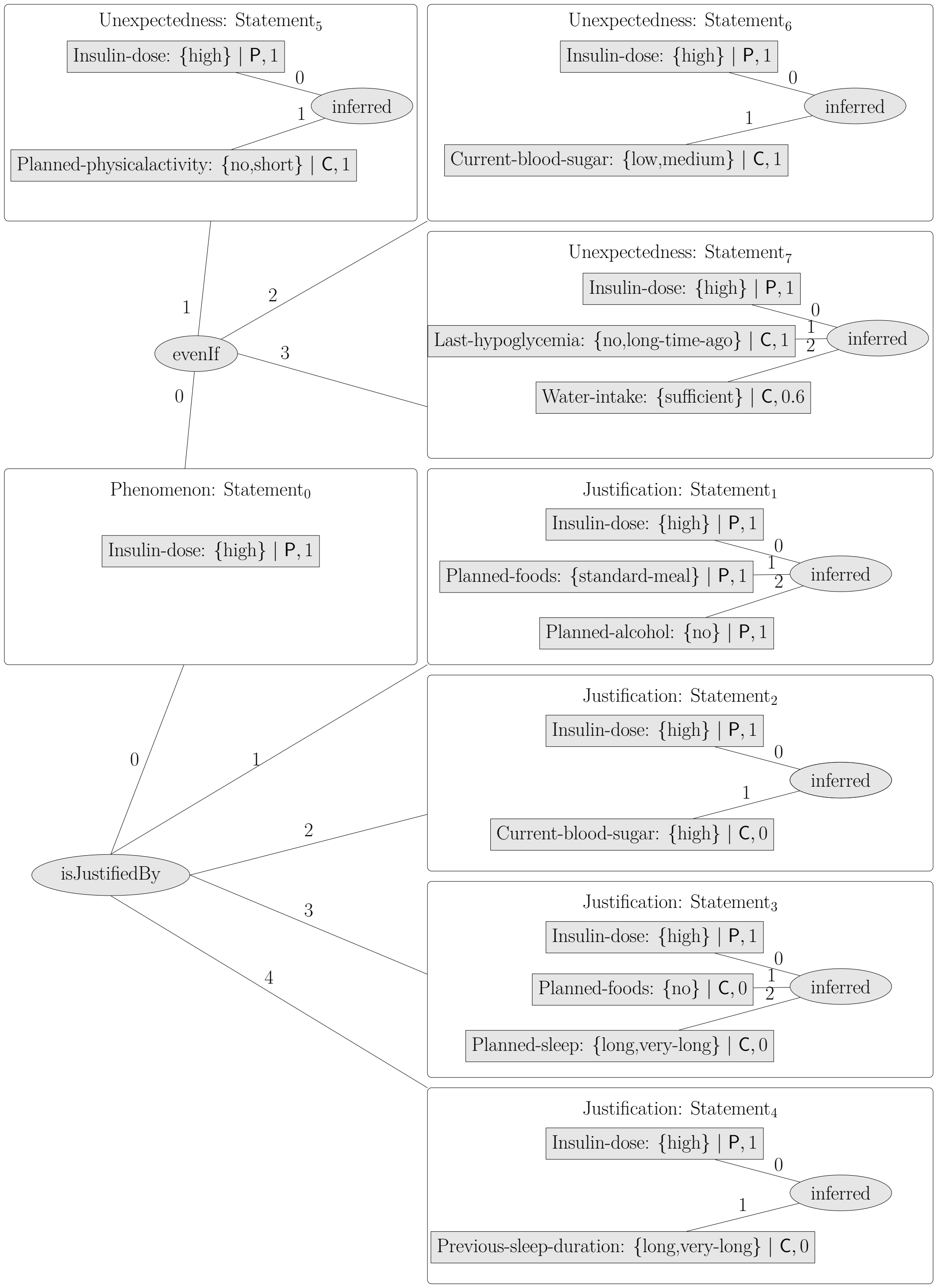


Figure 7.11: Representation of the combination of the justification and the unexpectedness of $\pi^*_{id(x)}(high)$.

7.7 Conclusion

In this chapter, we have graphically represented explanations of possibilistic inference decisions in terms of conceptual graphs. The methods for extracting the explanations were introduced in Chapter 4. The represented explanations are of three types: the justification of the possibility degree of an output attribute value, its unexpectedness, and a combination of the justification and the unexpectedness.
For representing the explanations, we used the framework presented in Chapter 6 that we extended by introducing *possibilistic conceptual graphs* to represent the statements of these explanations.We specified an input for the representation of an explanation, called a *possibilistic explanation query*, which allows us to determine the statements of the explanation and to construct the vocabulary associated to its representation. From this vocabulary, we built all the needed conceptual graphs of the representation. Then, by nesting the possibilistic conceptual graphs representing the statements in the root graph structuring the explanation, we obtained the representation of the explanation.

The representation of the combination of the justification and the unexpectedness of an output value is built similarly, but another root graph has to be defined, which links the phenomenon to the statements of the justification and those of the unexpectedness.

Chapter 8

Representation of explanations of fuzzy inference decisions

In this chapter, we represent graphically two explanations of the inference results of a Mamdani fuzzy inference system (a fuzzy rule-based system composed of possibility rules): the justification of a conclusion and its unexpectedness. The content of these explanations is extracted using the methods introduced in Chapter 5. As for the explanations of possibilistic inference decisions, we represent these explanations by conceptual graphs using the framework presented in Chapter 6. The constructions are similar to those of the explanations of possibilistic inference decisions that were represented in the previous chapter.
We start by setting the objects that compose an explanation of a fuzzy inference decision. For both explanations (justification and unexpectedness), the observed phenomenon is the fuzzy degree $\alpha^*_{(Z,O_j)}$ of the considered conclusion (Z, O_j). Depending on the explanation that is represented, the other statements are the fuzzy logic expressions (conjunction of propositions) that either compose the justification of the conclusion or its unexpectedness.
In the following, we define a *fuzzy conceptual graph* as a conceptual graph where each concept node is gifted with a fuzzy degree and a semantics (Section 8.1). Each statement of an explanation will be represented by a fuzzy conceptual graphs.
In Section 8.2, we specify the input of our representations, which we call a *fuzzy explanation query.* A fuzzy explanation query captures a justification or an unexpectedness, in order to establish the statements of the explanation.
In Section 8.3, from a fuzzy explanation query, we define a vocabulary that extends the one of the framework (Definition 6.1). Given this vocabulary, we construct fuzzy conceptual graphs representing the statements of an explanation and the root conceptual graph of the representation (Section 8.4). By nesting the fuzzy conceptual graphs representing the statements in the root conceptual graph (Definition 6.5), we obtain the representation of an explanation by a nested conceptual graph.

In Section 8.5, we represent an explanation that is a combination of the justification and the unexpectedness of a conclusion. This representation is obtained by combining the two fuzzy explanation queries associated respectively to the justification of a conclusion and to its unexpectedness into one and by using a new root graph to structure the explanation.
Finally, in Section 8.6, our constructions are illustrated by the explanations of the inference results of the Mamdani fuzzy inference system used as example in Chapter 5.

8.1 Fuzzy conceptual graphs

Fuzzy conceptual graphs were introduced by Morton [77] and appear in several works e.g. [30, 96, 103]. Compared to the previous approaches, we propose a definition of a fuzzy conceptual graph adapted to our needs, where we represent the semantics attached to a fuzzy degree of a concept node:

Definition 8.1 *A fuzzy conceptual graph (FCG) 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 $c \in C$:*

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

where $d_c \in [0, 1]$.

In our context of explanations of the inference results of a fuzzy rule-based system composed of possibility rules, we will only use the semantics P, for possible.
We naturally extend the definition of a *star BG* (Definition 1.8) i.e., a BG restricted to a relation node and its neighbors as a star FCG.
In a FCG, a concept node can represent an assertion of the form "X is A is possible to a degree d", where in its label, X is interpreted as a concept type, A as an individual marker, d as a degree and P is its semantics.

8.2 Fuzzy explanation query

To describe the vocabulary of an explanation, we introduce the notion of *fuzzy explanation query*:

Definition 8.2 *A fuzzy explanation query is formed by a triplet $\mathcal{E} = (\mathcal{T}, Z, O_j)$ such that:*
• *$\mathcal{T} = \{(p, sem, d)\}$ is a finite set of triplets (Notation 2),*

• *Z is the variable of an output linguistic variable z,*
• *$O_j \in T_z$ is a linguistic term of z such that $c = (Z, O_j)$ is a conclusion of the fuzzy rule-based system for which the justification or the unexpectedness of its fuzzy degree $\alpha^*_{(Z,O_j)}$ is requested.*

Let us set a fuzzy explanation query $\mathcal{E} = (\mathcal{T}, Z, O_j)$. To establish the settings of the explanation associated to $\mathcal{E}$ according to the framework for representing an explanation (Chapter 6), we state that the explanation is composed of $m + 1 = \text{card}(\mathcal{T}) + 1$ statements. The observed phenomenon is constructed using the variable Z, the linguistic term O_j and the fuzzy degree $\alpha^*_{(Z,O_j)}$. We adopt the following notations for $\mathcal{E}$:

Notation 4 *We index the triplets of $\mathcal{T}$ as follows:*

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

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 $j = 1, 2, \ldots, k_i$ we have:
• *$p_j^{(i)}$ is the fuzzy proposition $(X_j^{(i)}, A_j^{(i)})$, where $X_j^{(i)}$ is the variable of a linguistic variable $a_j^{(i)}$, $A_j^{(i)} \in T_{a_j^{(i)}}$ is one of its terms and $\alpha_{p_j^{(i)}}$ is its fuzzy degree.*
• *$\mathcal{X}^{(i)} = \{X_1^{(i)}, X_2^{(i)}, \ldots, X_{k_i}^{(i)}\}$ with $card(\mathcal{X}^{(i)}) = k_i$,*
• *$\mathcal{S}^{(i)} = \{A_1^{(i)}, A_2^{(i)}, \ldots, A_{k_i}^{(i)}\}$ with $card(\mathcal{S}^{(i)}) = k_i$.*

We take the *disjoint unions*:

$$\mathcal{X} = \dot{\bigcup}_{1\leq i\leq m} \mathcal{X}^{(i)} \text{ and } \mathcal{S} = \dot{\bigcup}_{1\leq i\leq m} \mathcal{S}^{(i)}.$$

From these disjoint unions, we can define an application τ: $\mathcal{S} \to \mathcal{X}$ such that $\tau(A_j^{(i)}) = X_j^{(i)}$. These disjoint unions are necessary because the terms $A_j^{(i)}$ and $A_{j'}^{(i')}$ associated to two variables $X_j^{(i)}$ and $X_{j'}^{(i')}$ with $i \neq i'$ in the two propositions $p_j^{(i)}$ and $p_{j'}^{(i')}$ may be equal.
Using the justification and the unexpectedness of a conclusion (Z, O_j), see equations (5.11) and (5.12), we form two distinct explanation queries:

$$\mathcal{E}_J = (\text{Justification}(Z, O_j), Z, O_j) \tag{8.1a}$$

and

$$\mathcal{E}_U = (\text{Unexpectedness}(Z, O_j), Z, O_j) \tag{8.1b}$$

8.3 Vocabulary construction

In this Section, we define a vocabulary $\mathcal{V}^{\mathcal{E}} = (T_C, T_R, \mathcal{I}, \tau, \sigma)$ associated to a fuzzy explanation query $\mathcal{E} = (\mathcal{T}, Z, O_j)$. Such vocabulary extends the vocabulary of the framework (Definition 6.1). In $\mathcal{V}^{\mathcal{E}}$, the variable Z and those in $\mathcal{X}$ are concept types. The linguistic term O_j and those 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 constructed from $\mathcal{V}^{\mathcal{E}}$ may contain:

- a concept node of type Z and individual marker O_j,
- a concept node of type $X_j^{(i)}$ and individual marker $A_j^{(i)}$, which gives a representation of the fuzzy 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 Z and the concept nodes representing the propositions $p_1^{(i)}, p_2^{(i)}, \ldots, p_{k_i}^{(i)}$.

The vocabulary $\mathcal{V}^{\mathcal{E}}$ includes the objects for constructing the root graph that structures the explanation. These objects are contained in the vocabulary of the framework (Definition 6.1).

Definition 8.3 *The vocabulary $\mathcal{V}^{\mathcal{E}}$ is defined by:*
• $T_C = \{Z\} \cup \mathcal{X} \cup \{\mathsf{e}\} \cup \{Phenomenon\}$ *with* $card(T_C) = 3 + \sum_{i=1}^{m} k_i$.
• $T_R = \{inferred_{v^{(i)}} | v^{(i)} \in \mathcal{T}\} \cup \{\mathsf{t}\}$ *with* $card(T_R) = m + 1$ *and such that* $arity(inferred_{v^{(i)}}) = k_i + 1$ *and* $arity(\mathsf{t}) = m + 1$.
• $\mathcal{I} = \{O_j\} \cup \mathcal{S} \cup \{Statement_0, Statement_1, \ldots, Statement_m\}$ *with* $card(\mathcal{I}) = m + 2 + \sum_{i=1}^{m} k_i$.
• $\tau : \mathcal{I} \to T_C$ *such that:*

- $O_j \mapsto Z$,
- $A_j^{(i)} \mapsto X_j^{(i)}$,
- $Statement_0 \mapsto Phenomenon$,
- $Statement_i \mapsto \mathsf{e}$ *for* $i = 1, 2, \ldots, m$.

• *The signature map σ is given by:*

- $\sigma(inferred_{v^{(i)}}) = (Z, X_1^{(i)}, X_2^{(i)}, \ldots, X_{k_i}^{(i)})$ *for* $v^{(i)} \in \mathcal{T}$,
- $\sigma(\mathsf{t}) = (Phenomenon, \mathsf{e}, \mathsf{e}, \ldots, \mathsf{e})$.

We form two vocabularies $\mathcal{V}^{\mathcal{E}_J}$ and $\mathcal{V}^{\mathcal{E}_U}$, which are associated to the two explanation queries $\mathcal{E}_J$ and $\mathcal{E}_U$, see (8.1).
In the vocabulary $\mathcal{V}^{\mathcal{E}_J}$, we note the concept type e: "Justification" and the relation symbol t: "isJustifiedBy". Similarly, in the vocabulary $\mathcal{V}^{\mathcal{E}_U}$, we note e and t "Unexpectedness" and "evenIf" respectively.

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

Let us specify, in a FCG $G = (C, R, E, l)$ built on the vocabulary $\mathcal{V}^{\mathcal{E}}$ (Definition 8.3), the definition of the labels of the following concept nodes:

- for a concept node $c \in C$ such that type$(c) = Z$ and marker$(c) = O_j$, we put:

$$sem_c = \mathsf{P} \text{ and } d_c = \alpha^*_{(Z,O_j)}. \tag{8.2}$$

- for a concept node $c \in C$ such that type$(c) = X_j^{(i)}$ and marker$(c) = A_j^{(i)}$, we take:

$$sem_c = sem^{(i)} \text{ and } d_c = \alpha_{(X_j^{(i)}, A_j^{(i)})}. \tag{8.3}$$

For the other concept nodes, we specify neither a degree nor a semantics. On the vocabulary $\mathcal{V}^{\mathcal{E}}$, let us define $m + 1$ FCG $D, N_1, N_2, \ldots, N_m$ as follows:

Definition 8.4 *D is defined as the FCG reduced to one concept node with label $(Z : O_j \mid \mathsf{P}, \alpha^*_{(Z,O_j)})$.*

Definition 8.5 *Each N_i is the star FCG 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: $c_Z^{(i)}, c_{X_1}^{(i)}, c_{X_2}^{(i)}, \ldots, c_{X_{k_i}}^{(i)}$ of type $Z, X_1^{(i)}, X_2^{(i)}, \ldots, X_{k_i}^{(i)}$ and marker $O_j, A_1^{(i)}, A_2^{(i)}, \ldots, A_{k_i}^{(i)}$, as in (8.2), (8.3). The multi-edges are labeled $(r_i, 0, c_Z^{(i)})$ and $(r_i, j, c_{X_j}^{(i)})$ for $j = 1, 2, \ldots, k_i$.*

To construct the graph R, we use Definition 6.4.
Finally, we use Definition 6.5 to obtain the representation of an explanation: the graphs $D, N_1, N_2, \ldots, N_m$ are nested in R.

8.5 Representation of the combination of the justification and the unexpectedness

Given a conclusion (Z, O_j) and the two fuzzy explanation queries $\mathcal{E}_J = (\mathcal{T}_J, Z, O_j)$ and $\mathcal{E}_U = (\mathcal{T}_U, Z, O_j)$ associated to the justification of the conclusion and its unexpectedness respectively, see (8.1), where $m_J = card(\mathcal{T}_J)$ and $m_U = card(\mathcal{T}_U)$, we can represent an explanation of $m+1 = m_J + m_U + 1$ statements, which is the combination of the two explanations. Such combination allows us to form new natural language explanations such as "even if p_1, Z is O_j is possible to a degree $\alpha^*_{(Z,O_j)}$ because of p_2", where $\{(p_1, sem_1, d_1)\}$ is an unexpectedness of a conclusion (Z, O_j) and $\{(p_2, sem_2, d_2)\}$ is its justification. To represent the combination, let us re-index the triplets in $\mathcal{T}_J$ from 1 to m_J, i.e., $\mathcal{T}_J = \{v^{(1)}, v^{(2)}, \ldots, v^{(m_J)}\}$ and those of $\mathcal{T}_U$ from $m_J + 1$ to $m = m_J + m_U$ i.e., $\mathcal{T}_U = \{v^{(m_J+1)}, v^{(m_J+2)}, \ldots, v^{(m)}\}$. We form a new fuzzy explanation query $\mathcal{E}_{JU}$ such that:

$$\mathcal{E}_{JU} = (\mathcal{T}_J \cup \mathcal{T}_U, Z, O_j) = (\mathcal{T}, Z, O_j), \tag{8.4}$$

where $\mathcal{T} = \mathcal{T}_J \cup \mathcal{T}_U$ is a disjoint union. From $\mathcal{E}_{JU}$, we adopt Notation 4 and obtain $\mathcal{X}$ and $\mathcal{S}$. We form a new vocabulary, which is slightly different from the one previously introduced (Definition 8.3). The concept type $\mathfrak{e}$ and the relation symbol $\mathfrak{t}$ are deleted. The new vocabulary contains two new concept types "Justification" and "Unexpectedness" and two new relation symbols: "isJustifiedBy" of arity $m_J + 1$ and "evenIf" of arity $m_U + 1$. The mappings of the vocabulary are updated. The individual typing function links statements to their respective types, and the signature map is given for the two new relations symbols.

Definition 8.6 *The vocabulary $\mathcal{V}^{\mathcal{E}_{JU}}$ is defined by:*
- $T_C = \{Z\} \cup \mathcal{X} \cup \{Phenomenon, Justification, Unexpectedness\}$ *with* $card(T_C) = 4 + \sum_{i=1}^{m} k_i$.
- $T_R = \{inferred_{v^{(i)}} | v^{(i)} \in \mathcal{T}\} \cup \{isJustifiedBy, evenIf\}$ *with* $card(T_R) = m{+}2$ *and such that* $arity(inferred_{v^{(i)}}) = k_i{+}1$, $arity(isJustifiedBy) = m_J{+}1$ *and* $arity(evenIf) = m_U + 1$.
- $\mathcal{I} = \{O_j\} \cup \mathcal{S} \cup \{Statement_0, Statement_1, \ldots, Statement_m\}$ *with* $card(\mathcal{I}) = m + 2 + \sum_{i=1}^{m} k_i$.
- $\tau : \mathcal{I} \to T_C$ *such that:*
  - $O_j \mapsto Z$,
  - $A_j^{(i)} \mapsto X_j^{(i)}$,
  - *Statement*$_0 \mapsto$ *Phenomenon,*
  - *Statement*$_i \mapsto$ *Justification for* $i = 1, 2, \ldots, m_J$.
  - *Statement*$_i \mapsto$ *Unexpectedness for* $i = m_J + 1, m_J + 2, \ldots, m = m_J + m_U$.
- *The signature map σ is given by:*
  - $\sigma(inferred_{v^{(i)}}) = (Z, X_1^{(i)}, X_2^{(i)}, \ldots, X_{k_i}^{(i)})$ *for* $v^{(i)} \in \mathcal{T}$,

- $\sigma(\textit{isJustifiedBy}) = (\textit{Phenomenon}, \textit{Justification}, \dots, \textit{Justification})$.
- $\sigma(\textit{evenIf}) = (\textit{Phenomenon}, \textit{Unexpectedness}, \dots, \textit{Unexpectedness})$.

From the vocabulary $\mathcal{V}^{\mathcal{E}_{JU}}$, we define the graphs of the representation. The FCG $D, N_1, N_2, \dots, N_m$ are constructed using Definitions 8.4 and 8.5. To build the root graph R, we use the definition of the root graph used for representing the combination of the justification and unexpectedness of the possibility degree of an output attribute value (Definition 7.7).
Finally, we use Definition 6.5 to obtain the representation of this explanation.

8.6 Examples

To illustrate our constructions, we rely on the fuzzy rule-based system used in the examples of Chapter 5. We remind that we extracted the justification of the three conclusions: $(Fbs, \textit{Low})$, $(Fbs, Medium)$ and $(Fbs, High)$ (Example 5.5) and the unexpectedness of $(Fbs, \textit{Low})$ (Example 5.6).
We start with the conclusion $(Fbs, Medium)$. To represent its justification, we formulate a fuzzy explanation query (Definition 8.1). Then, from it, we obtain a vocabulary, according to Definition 8.3:
• T_C = {Future-blood-sugar, Current-blood-sugar, Activity, Justification, Phenomenon}.
• T_R = {inferred, isJustifiedBy} such that:

$$\text{arity(inferred)} = 3 \quad \text{and} \quad \text{arity(isJustifiedBy)} = 2.$$

• $\mathcal{I} = \{\text{Medium}, \text{Eat}, \text{Low}, \text{Statement}_0, \text{Statement}_1\}$.
• $\tau : \mathcal{I} \to T_C$ such that:

- Medium $\mapsto$ Future-blood-sugar,
- Eat $\mapsto$ Activity,
- Low $\mapsto$ Current-blood-sugar,
- $\text{Statement}_0 \mapsto$ Phenomenon,
- $\text{Statement}_1 \mapsto$ Justification.

• The signature map σ is given by:

- $\sigma(\text{inferred}) = (\text{Future-blood-sugar}, \text{Activity}, \text{Current-blood-sugar})$
- $\sigma(\text{isJustifiedBy}) = (\text{Phenomenon}, \text{Justification})$.

We construct the graphs D, N_1 and R according to Definitions 8.4, 8.5, and 6.4) respectively, and obtain the representation of this justification (Figure 8.1) by nesting D and N_1 in R (Definition 6.5).

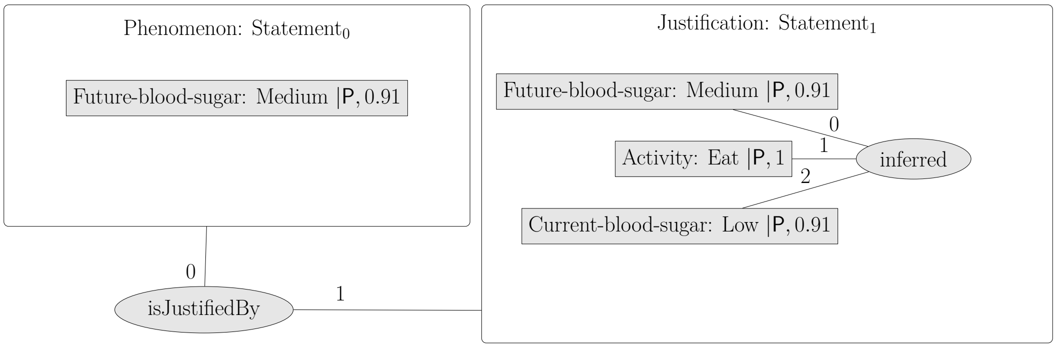


Figure 8.1: Representation of the justification of $(Fbs, Medium)$.

Similarly, we represent the justifications of $(Fbs, High)$ and (Fbs, Low), in Figures 8.2 and 8.3 respectively. For $(Fbs, High)$ the representation is composed of four graphs.

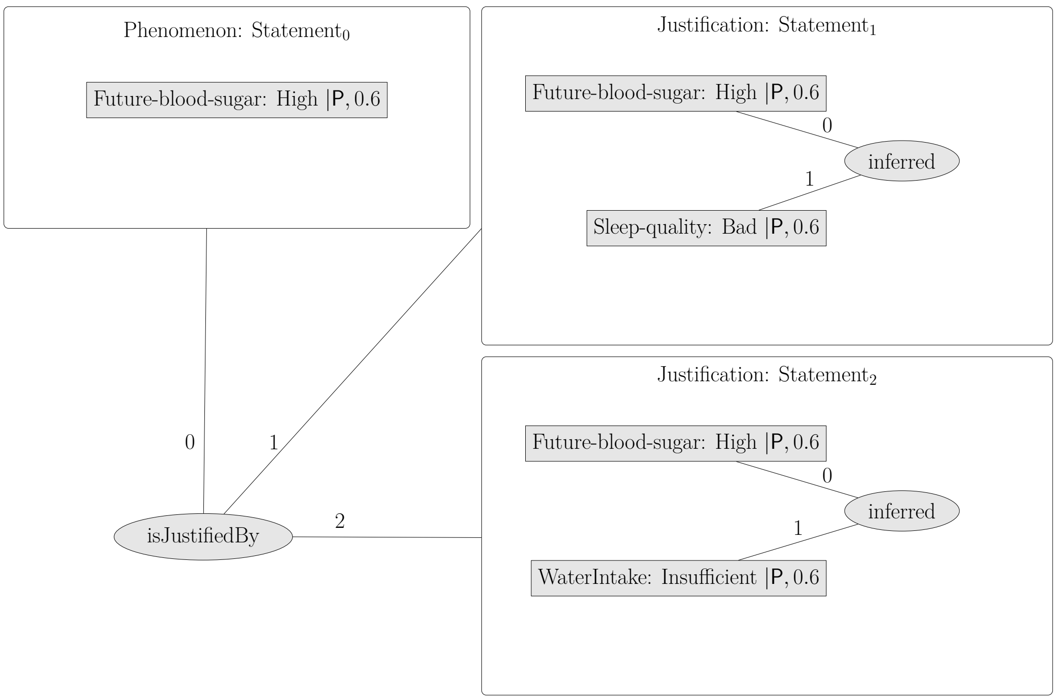


Figure 8.2: Representation of the justification of $(Fbs, High)$.

For representing the justification of (Fbs, Low), we construct seven graphs, as we extracted five triplets.

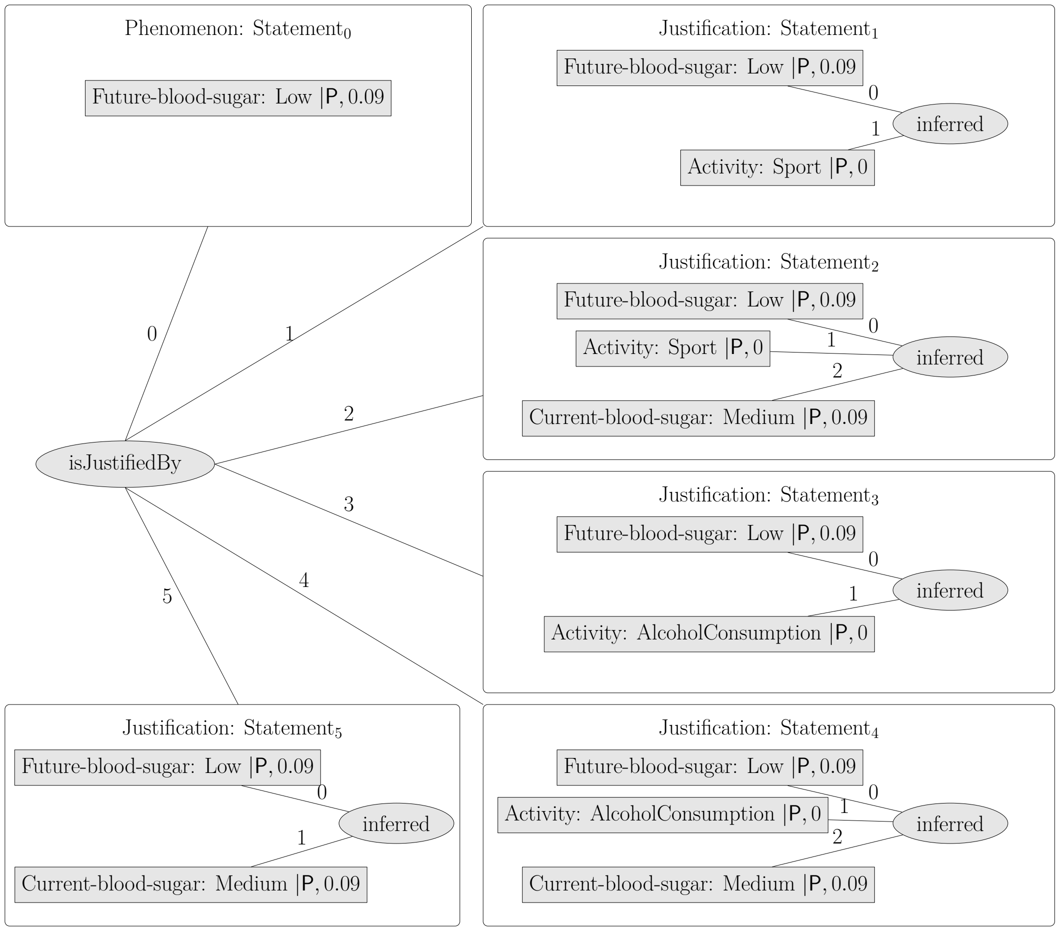


Figure 8.3: Representation of the justification of (Fbs, Low).

From the unexpectedness of (Fbs, Low), we form a corresponding fuzzy explanation query (Definition 8.1). It allows us to build an associated vocabulary, according to Definition 8.3. From this vocabulary, we construct the needed four graphs R, N_1, N_2, D to represent the explanation (Figure 8.4).

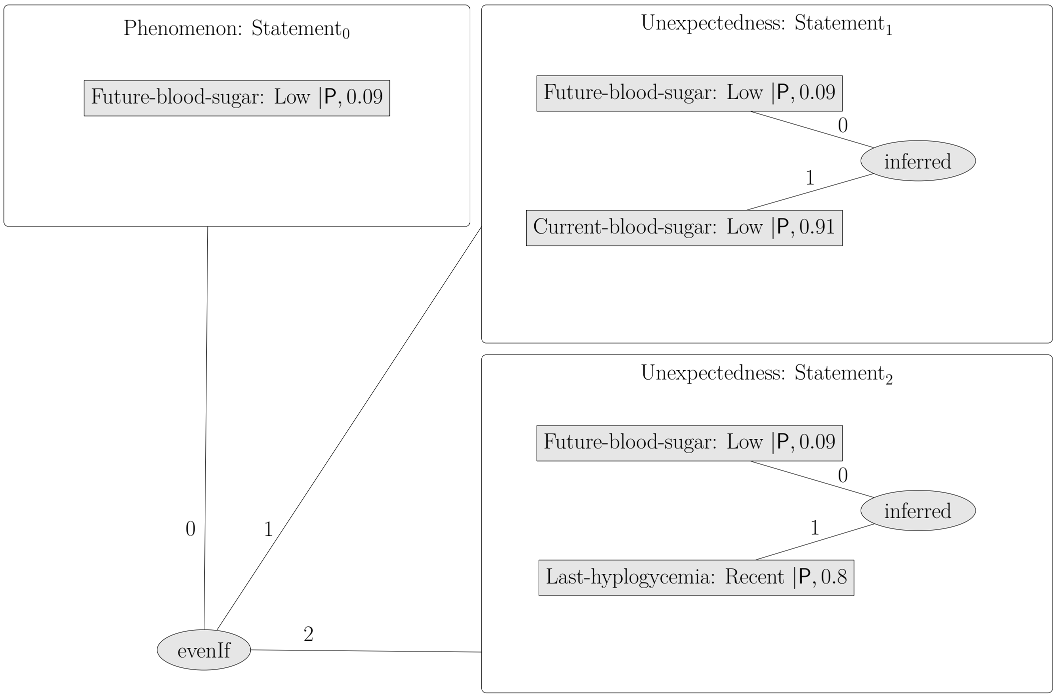


Figure 8.4: Representation of the unexpectedness of (Fbs, Low).

To represent the combination of the justification and the unexpectedness of (Fbs, Low), we combine the two associated fuzzy explanation queries into one, according to (8.4). Then, we obtain a new vocabulary (Definition 8.6). To represent the explanation, we construct another root graph R which contains two relation nodes and seven concept nodes (Definition 7.4). The conceptual graph R is represented in Figure 8.5.

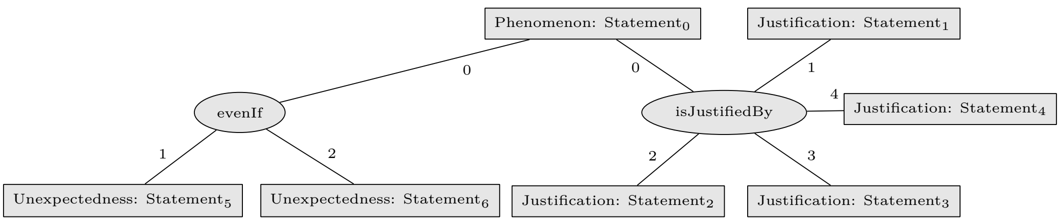


Figure 8.5: Graph R to represent the combination of the justification and the unexpectedness of (Fbs, Low).

The combination of the justification and the unexpectedness of (Fbs, Low) is represented in Figure 8.6.

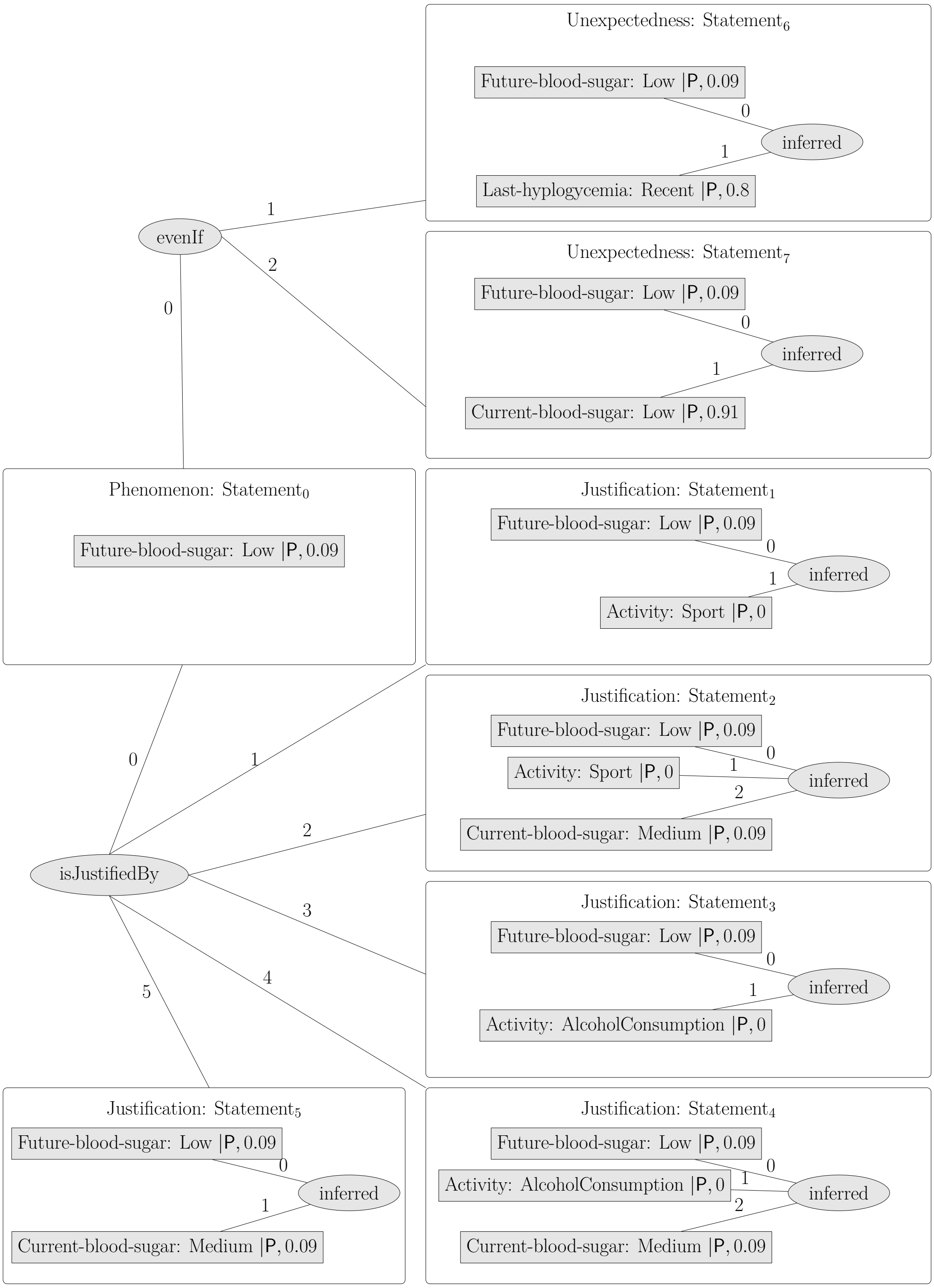


Figure 8.6: Representation of the combination of the justification and the unexpectedness of (Fbs, Low).

8.7 Conclusion

In this chapter, we represented three types of explanations of the inference results of a Mamdani fuzzy inference system in terms of conceptual graphs: the justification of a conclusion, its unexpectedness, and a combination of the justification and the unexpectedness. It is based on the methods developed in Chapter 5 for justifying the inference results of a Mamdani fuzzy inference system.

The representations of explanations of fuzzy inference decisions use the framework for representing the explanations developed in Chapter 6. We introduced a notion of a *fuzzy conceptual graph* that allows us to represent a fuzzy logic expression that is a conjunction of fuzzy propositions. We specified an input for the representation of an explanation, which is called a *fuzzy explanation query*. From a fuzzy explanation query, we built a vocabulary that allows us to construct graphs of the representation. The representation is obtained by nesting the graphs representing the statements in the root graph structuring the explanation, according to the framework (Chapter 6).

For representing the combination of the justification and unexpectedness of a conclusion, we combine their associated fuzzy explanation queries into one and form another root graph. Similarly as before, a vocabulary is formed, and it allows us to construct the graphs representing the statements that are then nested in the root graph.

Conclusion and Perspectives

In this thesis, we focused on two XAI objectives: the establishment of meeting points between KRR and ML and the elaboration of a processing chain for generating and evaluating AI explanations (Figure 1). Our explanatory paradigms were developed for two AI systems: a possibilistic rule-based system, where possibilistic rules encode negative information and a fuzzy rule-based system composed of possibility rules that encode positive information.

In Part B, for the first objective, we introduced a possibilistic interface between learning and if-then reasoning. The interface was defined by generalizing the min-max equation system of Farreny and Prade [55], which was proposed to develop the explanatory capabilities of possibilistic rule-based systems. From the generalized equation system, we obtained an explicit formula for the output possibility distribution, which allowed us to compute the corresponding possibility and necessity measures. We gave a necessary and sufficient condition for the output possibility distribution to be normalized and determined, when it is possible, minimal input solutions for the normalization. We have defined an algorithm to rebuild the equation system when we delete a rule. This algorithm allows us to obtain all equation subsystems of an initial equation system. Finally, we have shown that the equation system associated to a cascade can be represented by a min-max neural network.
From our generalized equation system, we may perform a sensitivity analysis, by setting the values of the input or output vector. This idea was originally suggested by Farreny and Prade [55]. To establish that a possibilistic rule base is coherent [50] i.e., given that the possibility distributions of the input attributes are normalized, the output possibility distribution must always be normalized, we may look for general conditions on the degrees of premises and parameters of rules. Finally, for developing possibilistic learning methods, it would be interesting to adapt, for our neural network, a min-max gradient descent method [22, 69, 95]. We may also consider using the learning method of the NEFLCASS model [79], which represents a fuzzy system with min-max common inference and uses a heuristic learning algorithm. Another approach for learning the parameters of the rules could be to take advantage of the fact that the equation system has the form of a system of fuzzy relation equations [86]. Therefore, the learning may be done by using the algorithms for the reso-

lution of fuzzy relation equations, see [82]. Approximation methods proposed in [31] may also be useful.

In Part C, for the processing chain, we introduced explanatory paradigms to justify the inference results of possibilistic and fuzzy rule-based systems. For both types of rule-based systems, we developed a method for selecting the rule premises that justify an inference result. Then, we defined premise reduction functions for both types of rule-based systems. By applying them to the selected premises, this allowed us to form two kind of explanations of an inference result: its justification and its unexpectedness. As our approach is based on a threshold that has a major impact on the content of the explanations, we need to find means to determine it for a rule base. This could perhaps be done in an incremental way.
It would also be important to evaluate the explanations to see if they are suitable for users [42, 75]. Evaluation protocols have been proposed for the explanations of the results of rule-based systems e.g., [15, 101]. We could also evaluate the impact of some of our explanatory methods on the user, for example, when does the user need the unexpectedness ? For possibilistic rule-based systems, does the proposition reduction methods bring benefits to the user?

In Part D, we proposed a graphical representation of an explanation. Firstly, we gave a general method for representing explanations in terms of conceptual graphs. Then, we extended it to represent explanations of possibilistic and fuzzy inference decisions. For each type of rule-based system, we represented three explanations of an inference result: its justification, its unexpectedness and a combination of its justification and its unexpectedness. For the representation of these explanations, we have defined two types of graphs: possibilistic conceptual graphs and fuzzy conceptual graphs, where each concept node is provided with a degree and an associated semantics. For our introduced graphs, we need to extend the work on classical conceptual graphs [34] to provide them with a query mechanism and a logical interpretation.
The representation can be extended for other explanations. For example, we may extend it to represent explanations of the inference results of a cascade (perhaps by nesting representations) or explanations of the results of other AI systems.
The representation may be used by natural language generation systems to produce natural language explanations. This could be done by adapting NLG systems that use semantic web inputs to produce text [24, 58]. Among them, note that the FORGe system [72], which obtained the highest score in the human evaluation of the WebNLG challenge [58], is based on the MATE graph transducer [23] that uses a conceptual graph as input.

Appendices

Appendix A

Publications

A.1 International Peer-Reviewed Conferences

- Baaj, I., Poli, J. P., Ouerdane, W., & Maudet, N. (2021, September). Representation of Explanations of Possibilistic Inference Decisions. In European Conference on Symbolic and Quantitative Approaches with Uncertainty (pp. 513-527). Springer, Cham.
- Baaj, I., Poli, J. P., Ouerdane, W., & Maudet, N. (2021, July). Min-max inference for Possibilistic Rule-Based System. In 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.
- Baaj, I., Poli, J. P., & Ouerdane, W. (2019). 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).
- Baaj, I., & Poli, J. P. (2019, June). Natural language generation of explanations of fuzzy inference decisions. In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.

A.2 National Peer-Reviewed Conferences

- Baaj, I., Poli, J. P., Ouerdane, W., & Maudet, N. (2021, October). Inférence min-max pour un système à base de règles possibilistes. In Rencontres francophones sur la logique floue et ses applications (pp. 233-240). Cepadues.

Appendix B

Programs

The repo:

`https://github.com/ibaaj/explainability-of-possibilistic-rule-based-systems`

contains three programs related to the paradigms presented in this thesis:

- The program in the “equation-system” folder allows us to construct the equation system associated to a cascade (Chapter 3).
- The program in the “example-1” folder allows us to form explanations of the inference results of the possibilistic rule-based system used as an example in both [17] and in Chapter 4.
- The program in the “example-2” folder allows us to form the explanations of the inference results of the second possibilistic rule-based system used as an example in Chapter 4.

The programs were tested with Python 3.9.8 and Mac OSX 11.5.2.

Bibliography

[1] US ACM. Public policy council.(2017). *Statement on algorithmic transparency and accountability*, 2017. (Cited on page 9.)

[2] José M Alonso, Ciro Castiello, Luis Magdalena, and Corrado Mencar. Design and validation of an explainable fuzzy beer style classifier. In *Explainable Fuzzy Systems*, pages 169–217. Springer, 2021. (Cited on page 39.)

[3] José M Alonso, Ciro Castiello, Luis Magdalena, and Corrado Mencar. Designing interpretable fuzzy systems. In *Explainable Fuzzy Systems*, pages 119–168. Springer, 2021. (Cited on page 39.)

[4] José M Alonso, Ciro Castiello, Luis Magdalena, and Corrado Mencar. *Explainable fuzzy systems: Paving the way from interpretable fuzzy systems to explainable AI systems*. Springer, 2021. (Cited on pages 9 and 39.)

[5] José M Alonso, Ciro Castiello, Luis Magdalena, and Corrado Mencar. An overview of fuzzy systems. *Explainable Fuzzy Systems*, pages 25–47, 2021. (Cited on pages 38 and 39.)

[6] José M Alonso, Luis Magdalena, and Serge Guillaume. Hilk: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *International Journal of Intelligent Systems*, 23(7):761–794, 2008. (Cited on page 39.)

[7] José M Alonso. From Zadeh's computing with words towards explainable artificial intelligence. In *International Workshop on Fuzzy Logic and Applications*, pages 244–248. Springer, 2018. (Cited on page 38.)

[8] José M Alonso and Alberto Bugarín. Expliclas: Automatic generation of explanations in natural language for weka classifiers. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2019. (Cited on page 40.)

[9] José M Alonso, Ciro Castiello, and Corrado Mencar. Interpretability of fuzzy systems: Current research trends and prospects. *Springer Handbook of Computational Intelligence*, pages 219–237, 2015. (Cited on page 39.)

[10] José M Alonso, Ciro Castiello, and Corrado Mencar. A bibliometric analysis of the explainable artificial intelligence research field. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 3–15. Springer, 2018. (Cited on page 39.)

[11] José M Alonso, Pietro Ducange, Riccardo Pecori, and Raúl Vilas. Building explanations for fuzzy decision trees with the expliclas software. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2020. (Cited on page 40.)

[12] José M Alonso, Alejandro Ramos-Soto, Ehud Reiter, and Kees van Deemter. An exploratory study on the benefits of using natural language for explaining fuzzy rule-based systems. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2017. (Cited on page 39.)

[13] Kay R Amel. From shallow to deep interactions between knowledge representation, reasoning and machine learning. In *Proceedings 13th International Conference on Scalable Uncertainty Management (SUM 2019), Compiegne, LNCS*, pages 16–18, 2019. (Cited on pages 10 and 42.)

[14] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020. (Cited on pages 9 and 10.)

[15] Ismaïl Baaj and Jean-Philippe Poli. Natural language generation of explanations of fuzzy inference decisions. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2019. (Cited on pages 96 and 144.)

[16] Ismaïl Baaj, Jean-Philippe Poli, and Wassila Ouerdane. 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)*, pages 14–19, 2019. (Cited on pages 7 and 10.)

[17] Ismaïl Baaj, Jean-Philippe Poli, Wassila Ouerdane, and Nicolas Maudet. Representation of explanations of possibilistic inference decisions. In *European Conference on Symbolic and Quantitative Approaches with Uncertainty*, pages 513–527. Springer, 2021. (Cited on page 149.)

[18] Jean-François Baget and Marie-Laure Mugnier. Extensions of simple conceptual graphs: the complexity of rules and constraints. *Journal of Artificial Intelligence Research*, 16:425–465, 2002. (Cited on page 31.)

[19] Regina Barzilay, Daryl McCullough, Owen Rambow, Jonathan DeCristofaro, Tanya Korelsky, and Benoit Lavoie. A new approach to expert system explanations. In *Natural Language Generation*, 1998. (Cited on page 38.)

[20] Salem Benferhat, Didier Dubois, and Henri Prade. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92(1-2):259–276, 1997. (Cited on page 81.)

[21] Or Biran and Courtenay Cotton. Explanation and justification in machine learning: A survey. In *IJCAI-17 workshop on explainable AI (XAI)*, volume 8, pages 8–13, 2017. (Cited on pages 9 and 10.)

[22] A Blanco, M Delgado, and I Requena. Solving fuzzy relational equations by max-min neural networks. In *Proceedings of 1994 IEEE 3rd International Fuzzy Systems Conference*, pages 1737–1742. IEEE, 1994. (Cited on page 143.)

[23] Bernd Bohnet and Leo Wanner. Open soucre graph transducer interpreter and grammar development environment. In *International Conference on Language Resources and Evaluation (LREC)*, 2010. (Cited on page 144.)

[24] Nadjet Bouayad-Agha, Gerard Casamayor, and Leo Wanner. Natural language generation in the context of the semantic web. *Semantic Web*, 5(6):493–513, 2014. (Cited on pages 11 and 144.)

[25] Bernadette Bouchon-Meunier, Didier Dubois, Lluis Godo, and Henri Prade. Fuzzy sets and possibility theory in approximate and plausible reasoning. In *Fuzzy sets in approximate reasoning and information systems*, pages 15–190. Springer, 1999. (Cited on pages 25, 27, 29, 30, 31, and 40.)

[26] Bernadette Bouchon-Meunier, Marie-Jeanne Lesot, and Christophe Marsala. Lotfi A. Zadeh, the visionary in explainable artificial intelligence. *TWMS Journal of Pure and Applied Mathematics*, 12(1):5–13, 2021. (Cited on page 38.)

[27] Adam Brown and Kelly L Close. *Bright spots & landmines: the diabetes guide I wish someone had handed me*. diaTribe Foundation, 2017. (Cited on pages 75, 91, 100, and 101.)

[28] James J Buckley and Yoichi Hayashi. Fuzzy neural networks: A survey. *Fuzzy sets and systems*, 66(1):1–13, 1994. (Cited on page 65.)

[29] Jean-Christophe Buisson, Henri Farreny, and Henri Prade. 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. (Cited on page 75.)

[30] Tru Hoang Cao. *Conceptual graphs and fuzzy logic: A fusion for representing and reasoning with linguistic information*, volume 306. Springer Science & Business Media, 2010. (Cited on page 132.)

[31] Katarína Cechlárová. A note on unsolvable systems of max–min (fuzzy) equations. *Linear Algebra and its Applications*, 310(1-3):123–128, 2000. (Cited on page 144.)

[32] Bruce Chandrasekaran, Michael C Tanner, and John R Josephson. Explaining control strategies in problem solving. *IEEE Intelligent Systems*, 4(01):9–15, 1989. (Cited on page 37.)

[33] Michel Chein and Marie-Laure Mugnier. Positive nested conceptual graphs. In *International Conference on Conceptual Structures*, pages 95–109. Springer, 1997. (Cited on page 33.)

[34] Michel Chein and Marie-Laure Mugnier. *Graph-based knowledge representation: computational foundations of conceptual graphs*. Springer Science & Business Media, 2008. (Cited on pages 11, 31, 34, 96, 110, 112, 116, and 144.)

[35] Michel Chein, Marie-Laure Mugnier, and Genevieve Simonet. Nested graphs: A graph-based knowledge representation model with fol semantics. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning*, pages 524–535, 1998. (Cited on pages 33 and 34.)

[36] William J Clancey and Reed Letsinger. *NEOMYCIN: Reconfiguring a rule-based expert system for application to teaching*. Department of Computer Science, Stanford University Stanford, CA, USA, 1982. (Cited on page 38.)

[37] Patricia Conde-Clemente, José M Alonso, and Gracian Trivino. rldcp: R package for text generation from data. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2017. (Cited on page 39.)

[38] Philip E Cryer. Hypoglycemia begets hypoglycemia in iddm. *Diabetes*, 42(12):1691–1693, 1993. (Cited on page 101.)

[39] Adnan Darwiche and Auguste Hirth. On the reasons behind decisions. In *24th European Conference on Artificial Intelligence (ECAI)*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 712–720. IOS Press, 2020. (Cited on page 90.)

[40] Jean-Louis Dessalles. Conversational topic connectedness predicted by simplicity theory. In *CogSci*, 2017. (Cited on page 11.)

[41] Judith P Dick. Using contexts to represent text. In *International Conference on Conceptual Structures*, pages 196–213. Springer, 1994. (Cited on page 33.)

[42] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017. (Cited on pages 10 and 144.)

[43] D. Dubois and H. Prade. *Possibility theory: an approach to computerized processing of uncertainty.* Plenum Press, New York, 1988. (Cited on pages 15 and 61.)

[44] D. Dubois and H. Prade. Possibility theory and its applications: Where do we stand? In *Handbook of Computational Intelligence*, 2015. (Cited on pages 15, 17, and 18.)

[45] Didier Dubois, Francesc Esteva, Lluís Godo, and Henri Prade. Fuzzy-set based logics-an history-oriented presentation of their main developments. *The Many Valued and Nonmonotonic Turn in Logic*, 8:325–449, 2007. (Cited on pages 9, 25, 26, 27, 29, and 30.)

[46] Didier Dubois, Petr Hájek, and Henri Prade. Knowledge-driven versus data-driven logics. *Journal of logic, Language and information*, 9(1):65–89, 2000. (Cited on page 18.)

[47] Didier Dubois and Henri Prade. Possibility theory as a basis for preference propagation in automated reasoning. In *IEEE International Conference on Fuzzy Systems*, pages 821–832. IEEE, 1992. (Cited on page 18.)

[48] Didier Dubois and Henri Prade. What are fuzzy rules and how to use them. *Fuzzy sets and systems*, 84(2):169–185, 1996. (Cited on page 23.)

[49] Didier Dubois and Henri Prade. Possibility theory: qualitative and quantitative aspects. In *Quantified representation of uncertainty and imprecision*, pages 169–226. Springer, 1998. (Cited on pages 21, 95, 97, and 98.)

[50] Didier Dubois and Henri Prade. From possibilistic rule-based systems to machine learning-a discussion paper. In *International Conference on Scalable Uncertainty Management*, pages 35–51. Springer, 2020. (Cited on pages 9, 11, 15, 41, 42, 43, 45, 54, 63, 65, and 143.)

[51] Didier Dubois, Henri Prade, and Laurent Ughetto. Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Transactions on Fuzzy Systems*, 5(3):398–417, 1997. (Cited on page 28.)

[52] Didier Dubois, Henri Prade, and Laurent Ughetto. A new perspective on reasoning with fuzzy rules. *International journal of intelligent systems*, 18(5):541–567, 2003. (Cited on pages 9, 23, and 24.)

[53] Henri Farreny and Henri Prade. Default and inexact reasoning with possibility degrees. *IEEE transactions on systems, man, and cybernetics*, 16(2):270–276, 1986. (Cited on pages 15, 40, and 76.)

[54] Henri Farreny and Henri Prade. Explications de raisonnements dans l'incertain. *Revue d'intelligence artificielle*, 4(2):43–75, 1990. (Cited on page 42.)

[55] Henri Farreny and Henri Prade. *Positive and Negative Explanations of Uncertain Reasoning in the Framework of Possibility Theory*, page 319–333. John Wiley & Sons, Inc., USA, 1992. (Cited on pages 9, 11, 41, 42, 43, 45, 53, 65, 66, 69, 71, 73, and 143.)

[56] Henri Farreny, Henri Prade, and E Wyss. Approximate reasoning in a rule-based expert system using possibility theory: A case study. In *IFIP Congress*, pages 407–414, 1986. (Cited on pages 15 and 40.)

[57] Alex A Freitas. Comprehensible classification models: a position paper. *ACM SIGKDD explorations newsletter*, 15(1):1–10, 2014. (Cited on page 9.)

[58] Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. The webnlg challenge: Generating text from rdf data. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 124–133, 2017. (Cited on pages 11 and 144.)

[59] Albert Gatt and Emiel Krahmer. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170, 2018. (Cited on page 10.)

[60] Albert Gatt and Ehud Reiter. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, pages 90–93, 2009. (Cited on page 40.)

[61] Bikash Chandra Ghosh and Vilas Wuwongse. Computational situation theory in the conceptual graph language. In *International Conference on Conceptual Structures*, pages 188–202. Springer, 1996. (Cited on page 33.)

[62] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), aug 2018. (Cited on page 10.)

[63] David Gunning and David Aha. DARPA's explainable artificial intelligence (xai) program. *AI Magazine*, 40(2):44–58, 2019. (Cited on page 9.)

[64] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. Xai—explainable artificial intelligence. *Science Robotics*, 4(37), 2019. (Cited on page 9.)

[65] Patrick Hayes. *RDF 1.1 semantics*. W3C Recommandation, 2004. Available on `https://www.w3.org/TR/rdf11-mt/` (accessed on October 12, 2021). (Cited on page 31.)

[66] Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4):e1312, 2019. (Cited on page 9.)

[67] Hazaël Jones, Brigitte Charnomordic, Didier Dubois, and Serge Guillaume. Practical inference with systems of gradual implicative rules. *IEEE Transactions on Fuzzy Systems*, 17(1):61–78, 2008. (Cited on pages 23, 24, 96, 97, 98, and 99.)

[68] Markus Langer, Daniel Oster, Timo Speith, Holger Hermanns, Lena Kästner, Eva Schmidt, Andreas Sesing, and Kevin Baum. What do we want from explainable artificial intelligence (xai)?–a stakeholder perspective on xai and a conceptual model guiding interdisciplinary xai research. *Artificial Intelligence*, 296:103473, 2021. (Cited on page 9.)

[69] Long Li, Zhijun Qiao, Yan Liu, and Yuan Chen. A convergent smoothing algorithm for training max–min fuzzy neural networks. *Neurocomputing*, 260:404–410, 2017. (Cited on page 143.)

[70] Zachary C Lipton. The mythos of model interpretability. *Queue*, 16(3):31–57, 2018. (Cited on page 9.)

[71] Ebrahim H Mamdani and Sedrak Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies*, 7(1):1–13, 1975. (Cited on page 26.)

[72] Simon Mille, Stamatia Dasiopoulou, and Leo Wanner. A portable grammar-based nlg system for verbalization of structured data. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, pages 1054–1056, 2019. (Cited on page 144.)

[73] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019. (Cited on page 9.)

[74] Brent Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2):2053951716679679, 2016. (Cited on page 9.)

[75] Sina Mohseni, Niloofar Zarei, and Eric D Ragan. A multidisciplinary survey and framework for design and evaluation of explainable ai systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3-4):1–45, 2021. (Cited on pages 10 and 144.)

[76] Claudio Moraga. An essay on the interpretability of mamdani systems. In *Combining Experimentation and Theory*, pages 61–72. Springer, 2012. (Cited on page 97.)

[77] S.K. Morton. *Conceptual graphs and fuzziness in artificial intelligence*. PhD thesis, University of Bristol, 1987. (Cited on page 132.)

[78] Shane T Mueller, Robert R Hoffman, William J Clancey, Abagail K Emery, and Gary Klein. Explanation in human-ai systems: A literature meta-review synopsis of key ideas and publications and bibliography for explainable ai. Technical report, DARPA, 2019. (Cited on pages 9 and 37.)

[79] Detlef Nauck and Rudolf Kruse. Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine*, 16(2):149–169, 1999. (Cited on page 143.)

[80] Adeline Nazarenko. Representing natural language causality in conceptual graphs: the higher order conceptual relation problem. In *International Conference on Conceptual Structures*, pages 205–222. Springer, 1993. (Cited on page 33.)

[81] Cécile L Paris, William R Swartout, and William C Mann. *Natural language generation in artificial intelligence and computational linguistics*, volume 119. Springer Science & Business Media, 2013. (Cited on page 38.)

[82] Ketty Peeva. Resolution of fuzzy relational equations–method, algorithm and software with applications. *Information Sciences*, 234:44–63, 2013. (Cited on page 144.)

[83] General Data Protection Regulation. Regulation eu 2016/679 of the european parliament and of the council of 27 april 2016. *Official Journal of the European Union*, 2016. (Cited on page 9.)

[84] Ehud Reiter and Robert Dale. Building applied natural language generation systems. *Natural Language Engineering*, 3(1):57–87, 1997. (Cited on pages 10 and 40.)

[85] Antoine Saillenfest and Jean-Louis Dessalles. Some probability judgments may rely on complexity assessments. In *CogSci*, 2015. (Cited on page 11.)

[86] Elie Sanchez. Resolution of composite fuzzy relation equations. *Information and Control*, 30(1):38–48, 1976. (Cited on page 143.)

[87] A Carlisle Scott, J William Clancey, Randall Davis, and Edward H Shortliffe. Methods for generating explanations. *Rule-Based Expert Systems, Addison-Wesley*, 1984. (Cited on page 38.)

[88] Edward H Shortliffe and Bruce G Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23(3-4):351–379, 1975. (Cited on page 37.)

[89] John F Sowa. Conceptual graphs for a data base interface. *IBM Journal of Research and Development*, 20(4):336–357, 1976. (Cited on page 31.)

[90] John F Sowa. Logical foundations for representing object-oriented systems. *Journal of Experimental & Theoretical Artificial Intelligence*, 5(2-3):237–261, 1993. (Cited on page 33.)

[91] Ilia Stepin, José M Alonso, Alejandro Catala, and Martin Pereira-Fariña. Generation and evaluation of factual and counterfactual explanations for decision trees and fuzzy rule-based classifiers. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2020. (Cited on page 40.)

[92] William R Swartout. Xplain: A system for creating and explaining expert consulting programs. *Artificial Intelligence*, 21(3):285–325, 1983. (Cited on page 38.)

[93] William R Swartout and Johanna D Moore. Explanation in second generation expert systems. In *Second generation expert systems*, pages 543–585. Springer, 1993. (Cited on page 38.)

[94] Randy L Teach and Edward H Shortliffe. An analysis of physician attitudes regarding computer-based clinical consultation systems. In *Use and impact of computers in clinical medicine*, pages 68–85. Springer, 1981. (Cited on pages 9 and 37.)

[95] Loo-Nin Teow and Kia-Fock Loe. An effective learning method for max-min neural networks. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 1134–1139. Citeseer, 1997. (Cited on page 143.)

[96] Rallou Thomopoulos, Patrice Buche, and Ollivier Haemmerlé. Representation of weakly structured imprecise data for fuzzy querying. *Fuzzy Sets and Systems*, 140(1):111–128, 2003. (Cited on page 132.)

[97] Nava Tintarev and Judith Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender systems Handbook*, pages 479–510. Springer, 2011. (Cited on page 10.)

[98] Gracian Trivino and Michio Sugeno. Towards linguistic descriptions of phenomena. *International Journal of Approximate Reasoning*, 54(1):22–34, 2013. (Cited on page 39.)

[99] Marie-Christine G Turnin, René H Beddok, Jacqueline P Clottes, Pascale F Martini, Robert G Abadie, Jean-Christophe Buisson, Chantal-Soulé Dupuy, Michel Bonneu, Rita Camaré, Jean-Paul Anton, et al. Telematic expert system diabeto: new tool for diet self-monitoring for diabetic patients. *Diabetes Care*, 15(2):204–212, 1992. (Cited on page 75.)

[100] Laurent Ughetto, Didier Dubois, and Henri Prade. Implicative and conjunctive fuzzy rules-a tool for reasoning from knowledge and examples. In *Association for the Advancement of Artificial Intelligence (AAAI)*, pages 214–219, 1999. (Cited on pages 15, 22, 23, 28, 96, 97, 98, 99, and 100.)

[101] Jasper van der Waa, Elisabeth Nieuwburg, Anita Cremers, and Mark Neerincx. Evaluating xai: A comparison of rule-based and example-based explanations. *Artificial Intelligence*, 291:103404, 2021. (Cited on pages 10 and 144.)

[102] Werner Van Leekwijck and Etienne E Kerre. Defuzzification: criteria and classification. *Fuzzy sets and systems*, 108(2):159–178, 1999. (Cited on pages 28 and 95.)

[103] Vilas Wuwongse and Mario Manzano. Fuzzy conceptual graphs. In *International Conference on Conceptual Structures*, pages 430–449. Springer, 1993. (Cited on page 132.)

[104] Victor L Yu, Lawrence M Fagan, Sharon M Wraith, William J Clancey, Carlisle A Scott, John Hannigan, Robert L Blum, Bruce G Buchanan, and Stanley N Cohen. Antimicrobial selection by a computer. a blinded evaluation by infectious diseases experts. *Jama*, 242(12):1279–1282, 1979. (Cited on page 37.)

[105] Lotfi A Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965. (Cited on pages 15 and 20.)

[106] Lotfi A Zadeh. The concept of a linguistic variable and its application to approximate reasoning—i. *Information Sciences*, 8(3):199–249, 1975. (Cited on page 22.)

[107] Lotfi A Zadeh. A theory of approximate reasoning (ar). Technical Report UCB/ERL M77/58, EECS Department, University of California, Berkeley, Aug 1977. (Cited on page 26.)

[108] Lotfi A Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1(1):3–28, 1978. (Cited on page 15.)

[109] Lotfi A Zadeh. A new direction in AI: Toward a computational theory of perceptions. *AI magazine*, 22(1):73–73, 2001. (Cited on page 39.)