



HAL
open science

Min-max inference for Possibilistic Rule-Based System

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

► **To cite this version:**

Ismail Baaj, Jean-Philippe Poli, Wassila Ouerdane, Nicolas Maudet. Min-max inference for Possibilistic Rule-Based System. 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Jul 2021, Luxembourg, Luxembourg. pp.9494506, 10.1109/FUZZ45933.2021.9494506 . hal-03408909

HAL Id: hal-03408909

<https://hal.science/hal-03408909v1>

Submitted on 19 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Min-max inference for Possibilistic Rule-Based System

Ismail Baaj,

Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France,
LIP6, Sorbonne Université, Paris, France,

Jean-Philippe Poli,

Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France,

Wassila Ouerdane, MICS, CentraleSupélec,
Université Paris-Saclay, Gif sur Yvette, France.

Nicolas Maudet,

LIP6, Sorbonne Université, Paris, France.

April 19, 2022

Abstract

In this paper, we explore the min-max inference mechanism of any rule-based system of n if-then possibilistic rules. We establish an additive formula for the output possibility distribution obtained by the inference. From this result, we deduce the corresponding possibility and necessity measures. Moreover, we give necessary and sufficient conditions for the normalization of the output possibility distribution.

As application of our results, we tackle the case of a cascade of two if-then possibilistic rules sets and establish an input-output relation between the two min-max equation systems. Finally, we associate to the cascade construction an explicit min-max neural network.

1 Introduction

Possibility Theory is a well-known framework for the representation of incomplete or imprecise information [1, 2] that has recently been emphasized as an interesting setting for the development of consistent interfaces between learning and reasoning [3]. Initially introduced by Zadeh [4] and considerably developed by Dubois and Prade, Possibility Theory is a natural complement of fuzzy theory which provides computable methods for modelling linguistically stated

information, but does not handle uncertainty, which is rather due to a lack of precise (or complete) information. Basically, in Possibility Theory, uncertainty is modelised by two dual measures called possibility and necessity, which allow to distinguish between what is possible without being certain at all and what is certain to some extent.

In the 80's, possibilistic handling of rule-based systems was developed [5, 6]. Then, in [7], the authors proposed an approach for the explanatory capabilities of possibilistic rule-based systems that relies on a min-max equation system. They used this equation system for performing a sensitivity analysis and illustrated it by an example of a system composed of three if-then possibilistic rules. These topics have recently been studied by [3] for developing possibilistic learning methods that would be consistent with if-then rule-based reasoning. The authors of [3] think that taking advantage of the min-max equation system of [7] may lead to the development of possibilistic learning methods. 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.

In this paper, we address a number of questions or issues mentioned or raised in [3, 7] by providing an in-depth study of the min-max equation system of a possibilistic rule-based system. We start by reminding the possibilistic rule-based systems and their underlying equation systems in section 2. In the case of n if-then possibilistic rules, we then give a canonical construction for the matrices of the equation system (section 3). This enables us to establish an additive formula for the output possibility distribution (section 4).

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 4). We also determine minimal input solutions for the normalization, when it is possible.

In section 5, 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 6) 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 7).

2 Background

Before introducing our work, we present a possibilistic rule-based system given by a set of if-then possibilistic rules. Then, we study the case of a cascade and remind the equation system introduced by [7]. Some notations are also given.

2.1 Possibilistic rule-based system

We consider a possibilistic rule-based system with a set of n if-then possibilistic rules R^1, R^2, \dots, R^n . Each R^i is a rule of the form: “if p_i then q_i ” and has its uncertainty propagation matrix $\begin{bmatrix} \pi(q_i|p_i) & \pi(q_i|\neg p_i) \\ \pi(\neg q_i|p_i) & \pi(\neg q_i|\neg p_i) \end{bmatrix} = \begin{bmatrix} 1 & s_i \\ r_i & 1 \end{bmatrix}$.

A non-compounded premise p_i is a proposition of the form “ $a_i(x) \in P_i$ ”. The attribute a_i is applied to an item x . Its information is represented by a possibility distribution $\pi_{a_i(x)} : D_{a_i} \rightarrow [0, 1]$ defined on its domain D_{a_i} , which is supposed to be normalized i.e., $\exists u \in D_{a_i}$ such that $\pi_{a_i(x)}(u) = 1$. The possibility degree of p_i and that of its negation are computed using the possibility measure Π by $\pi(p_i) = \Pi(P_i) = \sup_{u \in P_i} \pi_{a_i(x)}(u)$ and $\pi(\neg p_i) = \Pi(\overline{P_i}) = \sup_{u \in \overline{P_i}} \pi_{a_i(x)}(u)$ respectively, where P_i is a subset of D_{a_i} and $\overline{P_i}$ its complement. They are noted respectively λ_i and ρ_i . In the case of a compounded premise $p_i = p_{1,i} \wedge \dots \wedge p_{k,i}$, we take $\lambda_i = \min_{j=1}^k \pi(p_{j,i})$ and $\rho_i = \max_{j=1}^k \pi(\neg p_{j,i})$ [3].

The conclusion q_i represents “ $b(x) \in Q_i$ ”, where Q_i is a subset of D_b , the domain of the attribute b . The possibility degree of q_i and that of its negation are respectively noted α_i and β_i and are given by $\begin{bmatrix} \pi(q_i) \\ \pi(\neg q_i) \end{bmatrix} = \begin{bmatrix} 1 & s_i \\ r_i & 1 \end{bmatrix} \square_{\min}^{\max} \begin{bmatrix} \lambda_i \\ \rho_i \end{bmatrix}$. The operator \square_{\min}^{\max} uses min as the product and max as the addition. Taking into account the normalization of p_i i.e., $\max(\lambda_i, \rho_i) = 1$, we have $\alpha_i = \max(s_i, \lambda_i)$ and $\beta_i = \max(r_i, \rho_i)$. Finally, the possibility distribution of the attribute b associated with R^i is: $\pi_{b(x)}^{*i}(u) = \alpha_i \mu_{Q_i}(u) + \beta_i \mu_{\overline{Q_i}}(u)$ for any $u \in D_b$. With n rules, the possibility distribution of b is obtained by a min-based conjunctive combination:

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

2.2 Cascade

In this case, a possibilistic rule-based system relies on R^1, R^2, \dots, R^n and a new set of m if-then possibilistic rules R'^1, R'^2, \dots, R'^m , where the same attribute is used in both the conclusions of the R^i and the premises of the R'^j . 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 “ $c(x) \in Q''_c$ ” where Q''_c 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 .

2.3 Equation system

In [7], an equation system denoted $OV = MR \blacksquare IV$ was formulated, where OV and IV are respectively named the output and the input vectors. For an example of a system of three if-then possibilistic rules, the authors showed that the equation system describes the output possibility distribution and proposed

to use it to perform a sensitivity analysis. In [3], the equation system for the case of two rules R^1 and R^2 is given:

$$\begin{bmatrix} \Pi(Q_1 \cap Q_2) \\ \Pi(Q_1 \cap \overline{Q_2}) \\ \Pi(\overline{Q_1} \cap Q_2) \\ \Pi(\overline{Q_1} \cap \overline{Q_2}) \end{bmatrix} = \begin{bmatrix} s_1 & 1 & s_2 & 1 \\ s_1 & 1 & 1 & r_2 \\ 1 & r_1 & s_2 & 1 \\ 1 & r_1 & 1 & r_2 \end{bmatrix} \square_{\max}^{\min} \begin{bmatrix} \lambda_1 \\ \rho_1 \\ \lambda_2 \\ \rho_2 \end{bmatrix}. \quad (2)$$

The operator \square_{\max}^{\min} 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 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.

2.4 Notations

In our work, 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:

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

Finally, we introduce an operator denoted \square_{\min} that takes the minimum of the coefficients of each row in a matrix.

3 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, \dots, R^n . We introduce the generalized equation system, which we note:

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

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

3.1 Partition and settings

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

by the following two conditions:

$$\bullet E_1^{(1)} = Q_1 \text{ and } E_2^{(1)} = \overline{Q_1} \quad (3a)$$

and for $i > 1$:

$$\bullet E_k^{(i)} = \begin{cases} E_k^{(i-1)} \cap Q_i & \text{if } 1 \leq k \leq 2^{i-1} \\ E_{k-2^{i-1}}^{(i-1)} \cap \overline{Q_i} & \text{if } 2^{i-1} < k \leq 2^i \end{cases} \quad (3b)$$

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

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

3.2 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, \dots, N_{2^i} the rows of M_i .

3.3 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 $(2i, 1)$:

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

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

3.4 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} :

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

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

$$L_k = \begin{cases} (L'_k, \alpha_i) & \text{if } 1 \leq k \leq 2^{i-1} \\ (L'_{k-2^{i-1}}, \beta_i) & \text{if } 2^{i-1} < k \leq 2^i \end{cases}. \quad (4)$$

We can recover the matrix M_i from the matrix B_i in an explicit way.

3.5 Relation between B_i and the partition

For $k \in \{1, 2, \dots, 2^i\}$, let $L_k = (\gamma_1, \gamma_2, \dots, \gamma_i)$ 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:

$$E_k^{(i)} = T_1 \cap T_2 \cdots \cap T_i \text{ with } T_j = \begin{cases} Q_j & \text{if } \gamma_j = \alpha_j \\ \bar{Q}_j & \text{if } \gamma_j = \beta_j \end{cases}. \quad (5)$$

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 (4).

3.6 Coefficients of $\square_{\min} B_i$

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

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

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

$$o_k^{(i)} = \begin{cases} \min(o_k^{(i-1)}, \alpha_i) & \text{if } 1 \leq k \leq 2^{i-1} \\ \min(o_{k-2^{i-1}}^{(i-1)}, \beta_i) & \text{if } 2^{i-1} < k \leq 2^i \end{cases}. \quad (6)$$

This result is directly deduced from (4) and the associativity of the min function. Finally, we obtain:

Theorem 1 *The min-max matrix product of M_i by the matrix-column I_i is obtained by applying the operator \square_{\min} to the matrix B_i :*

$$M_i \square_{\max}^{\min} I_i = \square_{\min} B_i \quad (7)$$

The abridged proof of this result goes as follows. It is clear for $i = 1$. Starting from the assumption (7) for $i - 1$, we show that each coefficient $o_k^{(i)}$ of $\square_{\min} B_i$ is equal to $N_k \square_{\max}^{\min} I_i$, N_k being the k -th row of M_i .

4 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 $(E_k^{(i)})_{1 \leq k \leq 2^i}$. 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.

4.1 Additive formula for $\pi_{b(x)}^*$

We use the coefficients of $O_i = \square_{\min} B_i$, see (7), and introduce the characteristic functions $\mu_{E_1^{(i)}}, \mu_{E_2^{(i)}}, \dots, \mu_{E_{2^i}^{(i)}}$ of the sets $E_1^{(i)}, E_2^{(i)}, \dots, E_{2^i}^{(i)}$.

Theorem 2 *The output possibility distribution $\pi_{b(x),i}^*$ associated to the first i rules is:*

$$\pi_{b(x),i}^* = \sum_{1 \leq k \leq 2^i} o_k^{(i)} \mu_{E_k^{(i)}}. \quad (8)$$

We prove this result by recurrence on i , using (3), (6), and the associativity of the min function.

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

$$\exists k \in \{1, 2, \dots, 2^i\} \text{ such that } E_k^{(i)} \neq \emptyset \text{ and } o_k^{(i)} = 1. \quad (9)$$

Moreover, from (8), we deduce that the possibility measure of each *non-empty* set $E_k^{(i)}$ of the partition is equal to $o_k^{(i)}$. It is then natural to introduce:

$$J = \left\{ k \in \{1, 2, \dots, 2^i\} \mid E_k^{(i)} \neq \emptyset \right\} \text{ and } \omega = \text{card}(J).$$

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:

$$[\Pi(E_k^{(i)})]_{k \in J} = [o_k^{(i)}]_{k \in J}.$$

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 .

4.2 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 (7) and (8), we have $\pi_{b(x),i}^*(u) = N \square_{\max}^{\min} I_i$, where N is a row of \mathcal{M}_i . Let us note:

$$N = [t_1, t_2, \dots, t_{2i-1}, t_{2i}] \text{ where } t_j \in \{1, r_*, s_*\}.$$

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 $E = \{j \in \{1, 2, \dots, 2i\} \mid t_j < 1\}$. 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 [7].

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: $\forall k \in \{1, 2, \dots, i\}: \max(x_{2k-1}, x_{2k}) = 1$, that satisfies the following equation:

$$N \square_{\max}^{\min} X = 1. \quad (10)$$

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

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

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 $j = 1, 2, \dots, 2i$. For this maximal solution S_{\max} , we notice that $\forall k \in \{1, 2, \dots, i\}, (\lambda_k, \rho_k) = (1, 1)$: for each rule, the premise is considered as unknown [3].

Let us look for a unique minimal solution:

- If we have:

$$\exists k \in \{1, 2, \dots, i\} \text{ such that } (t_{2k-1}, t_{2k}) = (1, 1), \quad (11)$$

then equation (10) does not admit a minimal solution.

- If we suppose that:

$$\forall k \in \{1, 2, \dots, i\} \text{ we have } (t_{2k-1}, t_{2k}) \neq (1, 1), \quad (12)$$

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

4.3 Rebuild the equation system with a deleted rule

Let us delete a rule R^z from R^1, R^2, \dots, R^i and study the corresponding equation system. We denote by B_i^z the matrix associated to this system of $i - 1$

rules by the construction of subsection D of 3. We assert that the matrix B_i^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^z . Then, one can deduce the partition $(E_k^z)_{1 \leq k \leq 2^{i-1}}$ of D_b from B_i^z by using the relation (5). Its sets can also be determined directly from the partition $(E_k^{(i)})_{1 \leq k \leq 2^i}$. In fact, for any $k \in \{1, 2, \dots, 2^{i-1}\}$, we can find two indices $k', k'' \in \{1, 2, \dots, 2^i\}$ such that:

$$E_k^z = E_{k'}^{(i)} \cup E_{k''}^{(i)} \quad (13)$$

where, with respect to the notations of (5), $E_{k'}^{(i)}$ and $E_{k''}^{(i)}$ differ only on the component T_z , e.g. $T_z = Q_z$ for $E_{k'}^{(i)}$, and $T_z = \overline{Q}_z$ for $E_{k''}^{(i)}$. If $E_k^z \neq \emptyset$, the decomposition (13) is unique, where $E_{k'}^{(i)}$ and $E_{k''}^{(i)}$ still satisfy the above assumption. These two sets can be easily determined algorithmically.

Finally, we deduce from (13), that $\omega^z \leq \omega$, where $\omega^z = \text{card}(J^z)$ and $J^z = \{k \in \{1, 2, \dots, 2^{i-1}\} \mid E_k^z \neq \emptyset\}$. The partition $(E_k^z)_{k \in J^z}$ and the matrix B_i^z are particularly interesting for performing a sensitivity analysis. We illustrate the construction of B_i^z for the case $i = 3$ and $z = 2$:

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \alpha_2 & \alpha_3 \\ \alpha_1 & \beta_2 & \alpha_3 \\ \beta_1 & \beta_2 & \alpha_3 \\ \alpha_1 & \alpha_2 & \beta_3 \\ \beta_1 & \alpha_2 & \beta_3 \\ \alpha_1 & \beta_2 & \beta_3 \\ \beta_1 & \beta_2 & \beta_3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} \alpha_1 & \alpha_3 \\ \beta_1 & \alpha_3 \\ \alpha_1 & \alpha_3 \\ \beta_1 & \alpha_3 \\ \alpha_1 & \beta_3 \\ \beta_1 & \beta_3 \\ \alpha_1 & \beta_3 \\ \beta_1 & \beta_3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} \alpha_1 & \alpha_3 \\ \beta_1 & \alpha_3 \\ \alpha_1 & \beta_3 \\ \beta_1 & \beta_3 \end{bmatrix} = B_3^z.$$

5 Measures of possibility and necessity 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 [1]:

$$\Pi^* : \mathcal{P}(D_b) \rightarrow [0, 1] : Q \mapsto \Pi^*(Q) = \max_{u \in Q} \pi_{b(x)}^*(u),$$

$$N^* : \mathcal{P}(D_b) \rightarrow [0, 1] : Q \mapsto N^*(Q) = 1 - \Pi^*(\overline{Q}).$$

For a proposition p : “ $b(x) \in Q$ ”, the possibility degree $\pi(p)$ and its degree of necessity $n(p)$ are given by [1]:

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

In what follows, we deduce explicit formulas for $\Pi^*(Q)$ and $N^*(Q)$ from (8), where we use a function ε , which checks if a set is not empty:

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

5.1 Possibility measure

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

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

Therefore we can restate this result as:

$$\begin{aligned} \Pi^*(Q) &= \max_{1 \leq i \leq \omega} \varepsilon(E_{k_i}^{(n)} \cap Q) \cdot o_{k_i}^{(n)} \\ &= \max_{1 \leq i \leq \omega} \min(\varepsilon(E_{k_i}^{(n)} \cap Q), o_{k_i}^{(n)}). \end{aligned} \quad (14)$$

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

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

Then, equality (14) is exactly:

$$\Pi^*(Q) = \Delta_Q \square_{\min}^{\max} \mathcal{O}_n. \quad (15)$$

5.2 Necessity measure

Using (15), we have for \overline{Q} :

$$\Pi^*(\overline{Q}) = \Delta_{\overline{Q}} \square_{\min}^{\max} \mathcal{O}_n.$$

The necessity measure is then:

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

By the correspondences between \square_{\max}^{\min} and \square_{\min}^{\max} we obtain:

$$N^*(Q) = (\Delta_{\overline{Q}} \square_{\min}^{\max} \mathcal{O}_n)^\circ = \Delta_{\overline{Q}}^\circ \square_{\max}^{\min} \mathcal{O}_n^\circ. \quad (16)$$

6 Cascade

In this section, we use two sets of if-then possibilistic rules: the n rules R^1, R^2, \dots, R^n and the m rules R'^1, R'^2, \dots, R'^m . We form $\mathcal{O}_n = \mathcal{M}_n \square_{\max}^{\min} I_n$ for the first set of rules and $\mathcal{O}'_m = \mathcal{M}'_m \square_{\max}^{\min} I'_m$ for the second one, where we consider that their associated partition of non-empty sets $(E_k^{(n)})_{k \in J}$ and $(E_k'^{(m)})_{k \in J'}$ 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.

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

$$I'_m = \Delta \square_{\min}^{\max} \mathcal{O}_n. \quad (17)$$

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

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

We note S_1, S_2, \dots, S_{2m} the rows of Δ . Each row of Δ is in fact $\Delta_{Q'}$ associated to some set $Q' \subseteq D_b$ as in (15):

$$\Delta = \begin{bmatrix} \Delta_{Q'_1} \\ \Delta_{\overline{Q'_1}} \\ \vdots \\ \Delta_{Q'_m} \\ \Delta_{\overline{Q'_m}} \end{bmatrix}.$$

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, Δ and \mathcal{M}'_m :

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

6.2 Representation by a min-max neural network

In [3], 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° , \mathcal{M}_n° and $\mathcal{M}'_m{}^\circ$, we can express the equations involved in the cascade using only the operator $(A \square_{\min}^{\max} B)^\circ$: $\mathcal{O}_n = (\mathcal{M}_n^\circ \square_{\min}^{\max} I_n^\circ)^\circ$, $I'_m{}^\circ = (\Delta \square_{\min}^{\max} \mathcal{O}_n)^\circ$ 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, \dots, i_{2n} with $z_{i_1}, z_{i_2}, \dots, z_{i_{2n}}$ being their respective output values,
- the layer 2 has ω hidden neurons: $h_1, h_2, \dots, h_\omega$ where $I_{h_1}, I_{h_2}, \dots, I_{h_\omega}$ are their respective input values and $z_{h_1}, z_{h_2}, \dots, z_{h_\omega}$ their respective output values,
- the layer 3 has $2m$ hidden neurons: $h'_1, h'_2, \dots, h'_{2m}$ where $I_{h'_1}, I_{h'_2}, \dots, I_{h'_{2m}}$ are their respective input values and $z_{h'_1}, z_{h'_2}, \dots, z_{h'_{2m}}$ their respective output values,
- the layer 4 has ω' output neurons: $o'_1, o'_2, \dots, o'_{\omega'}$ where $I_{o'_1}, I_{o'_2}, \dots, I_{o'_{\omega'}}$ are their respective input values and $z_{o'_1}, z_{o'_2}, \dots, z_{o'_{\omega'}}$ 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 $f(x) = x$.

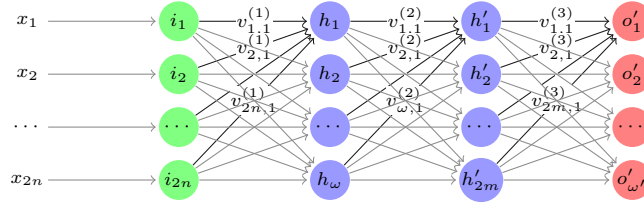


Figure 1: Min-max neural network architecture

We explicit its architecture (Figure 1) and define the following edges by:

- $x_j = 1 - \theta_j$, a coefficient of I_n° with $1 \leq j \leq 2n$,
- $v_{i,j}^{(1)} = 1 - t_{j,i}$, a coefficient of \mathcal{M}_n° with $1 \leq j \leq \omega$ and $1 \leq i \leq 2n$,
- $v_{i,j}^{(2)} = \delta_{j,i}$, a coefficient of Δ with $1 \leq j \leq 2m$ and $1 \leq i \leq \omega$,
- $v_{i,j}^{(3)} = 1 - t'_{j,i}$, 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 :

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

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

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

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 :

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

We get the output value of o'_k with $z_{o'_k} = f(I_{o'_k}) = I_{o'_k}$. So $z_{o'_1}, z_{o'_2}, \dots, z_{o'_\omega}$ are the coefficients of \mathcal{O}'_m .

As characteristics, we notice that each edge $v_{i,j}^{(2)}$ has a value equal to 0 or 1 with respect to the relation (17), while the values of the others are in $[0, 1]$. Furthermore, it has some resemblance with an hybrid fuzzy neural network [8], 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.

6.3 Example

To illustrate the cascade, we use the example of [3], previously introduced in the french version of [7]. It is a possibilistic rule-based system which suggests to people professions with an 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. For this set of rules, an equation system is formed, where \mathcal{O}_n has five coefficients. In fact, the height possible professions are in five non-empty disjoint sets which form a partition: $E_{k_1}^{(3)} = \{\text{researcher}\}$, $E_{k_2}^{(3)} = \{\text{professor}\}$, $E_{k_3}^{(3)} = \{\text{engineer, architect}\}$, $E_{k_4}^{(3)} = \{\text{business man, lawyer, doctor}\}$ and $E_{k_5}^{(3)} = \{\text{others}\}$. With the possibility distributions of the input attributes of [7], we get: $\Pi(E_{k_1}^{(3)}) = 0.2$, $\Pi(E_{k_2}^{(3)}) = 1$, $\Pi(E_{k_3}^{(3)}) = 0.2$, $\Pi(E_{k_4}^{(3)}) = 0.6$ and $\Pi(E_{k_5}^{(3)}) = 0.5$.

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 rather low,

- R'^2 : if a person is an engineer, a lawyer or an architect, 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, the sets $E'_{k_1(3)} = \{\text{high}\}$, $E'_{k_2(3)} = \{\text{average}\}$ and $E'_{k_3(3)} = \{\text{low}\}$ 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 Δ and then I'_m :

$$\Delta = \begin{bmatrix} 11000 \\ 00111 \\ 00110 \\ 11011 \\ 00010 \\ 11111 \end{bmatrix} \quad \text{and} \quad I'_m = \begin{bmatrix} 1 \\ 0.6 \\ 0.6 \\ 1 \\ 0.6 \\ 1 \end{bmatrix}.$$

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 $\mathcal{O}'_m = \mathcal{M}'_m \square_{\max}^{\min} I'_m$:

$$\begin{bmatrix} \Pi(E'_{k_1(3)}) \\ \Pi(E'_{k_2(3)}) \\ \Pi(E'_{k_3(3)}) \end{bmatrix} = \begin{bmatrix} 1 & 0.7 & 0.8 & 1 & 0.6 & 1 \\ 1 & 0.7 & 0.8 & 1 & 1 & 0.4 \\ 1 & 1 & 1 & 0.2 & 1 & 0.4 \end{bmatrix} \square_{\max}^{\min} \begin{bmatrix} 1 \\ 0.6 \\ 0.6 \\ 1 \\ 0.6 \\ 1 \end{bmatrix}.$$

and get: $\Pi(E'_{k_1(3)}) = 0.6$, $\Pi(E'_{k_2(3)}) = 0.7$ and $\Pi(E'_{k_3(3)}) = 1$. Such cascade is represented by a min-max neural network (see 6.2).

7 Conclusion

In this paper, we gave a canonical construction for the matrices governing a min-max equation system associated to a possibilistic rule-based system. As consequences of this construction, we obtained an additive formula for the output possibility distribution that allowed us to tackle the normalization problem of it and the case of a cascade. We showed that the equation system for a cascade is a nesting of max-min products of matrices and represented such construction by a min-max neural network. Therefore, the cascade is represented by a graph, which offers interesting perspectives for tasks such as sensitivity analysis. This possibilistic neuro-symbolic method allows to glimpse how possibilistic learning can have a consistent correspondence with possibilistic rule-based systems.

References

- [1] D. Dubois and H. Prade, *Possibility theory : an approach to computerized processing of uncertainty*. New York: Plenum Press, 1988.

- [2] —, “Possibility theory and its applications: Where do we stand?” in *Springer Handbook of Computational Intelligence*, J. Kacprzyk and W. Pedrycz, Eds. Springer Berlin Heidelberg, 2015, pp. 31–60.
- [3] —, “From possibilistic rule-based systems to machine learning - a discussion paper,” in *Scalable Uncertainty Management*, J. Davis and K. Tabia, Eds. Cham: Springer International Publishing, 2020, pp. 35–51.
- [4] L. A. Zadeh, “Fuzzy sets as a basis for a theory of possibility,” *Fuzzy sets and systems*, vol. 1, no. 1, pp. 3–28, 1978.
- [5] H. Farreny and H. Prade, “Default and inexact reasoning with possibility degrees,” *IEEE transactions on systems, man, and cybernetics*, vol. 16, no. 2, pp. 270–276, 1986.
- [6] H. Farreny, H. Prade, and E. Wyss, “Approximate reasoning in a rule-based expert system using possibility theory: A case study.” in *IFIP Congress*, 1986, pp. 407–414.
- [7] H. Farreny and H. Prade, “Positive and negative explanations of uncertain reasoning in the framework of possibility theory,” in *Fuzzy logic for the management of uncertainty*, 1992, pp. 319–333.
- [8] J. J. Buckley and Y. Hayashi, “Fuzzy neural networks: A survey,” *Fuzzy sets and systems*, vol. 66, no. 1, pp. 1–13, 1994.