# Uncertainty and explanations
# - Possibilistic and analogical settings -

Henri Prade

IRIT - CNRS, 118, route de Narbonne, 31062 Toulouse Cedex 09, France
prade@irit.fr

**Abstract.** This note briefly presents two different types of approaches to explanations, one that originates in the setting of possibility theory and the other that relies on the use of analogical proportions. The ideas are only outlined, but the main references are provided.

## 1 Introduction

The first trend of researches on explanations in AI took place in the late 1980's and early 1990's at the end of the boom on expert systems, see [6] for a an overview. At that time, expert knowledge was often considered as being pervaded with uncertainty, and this led to a few works on explanations in presence of uncertainty, in particular represented in the framework of belief functions [13], or in the framework of possibility theory [8].

In the following, we briefly revisit this latter work, before considering the analogical reasoning setting where conclusions are brittle and uncertain.

## 2 Possibilistic setting

In possibility theory, there exists a matrix calculus (which can be closely related to possibilistic logic), that handles uncertain rules "if $p$ then $q$", rather than undirected logical formulas [8, 7]. It relies on the following $\max - \min$ matrix product (here denoted $\otimes$):

$$\begin{bmatrix} \pi(q) \\ \pi(\neg q) \end{bmatrix} = \begin{bmatrix} \pi(q|p) & \pi(q|\neg p) \\ \pi(\neg q|p) & \pi(\neg q|\neg p) \end{bmatrix} \otimes \begin{bmatrix} \pi(p) \\ \pi(\neg p) \end{bmatrix}.$$

where the conditional possibility distributions obey a qualitative form of conditioning.

Due to normalization conditions (e.g., $\max(\pi(p), \pi(\neg p)) = 1$), which are preserved by the product $\otimes$, a matrix $\begin{bmatrix} 1 & s \\ r & 1 \end{bmatrix}$ represents the rules "if $p$ then $q$" (with certainty $1 - r$) and "if $\neg p$ then $\neg q$" (with certainty $1 - r$).

Let us consider a set of $m$ parallel uncertain rules of the form "if $a_i^1(x)$ is $P_i^1$ and $\cdots$ and $a_i^1(x)$ is $P_i^k$ then $b_i(x)$ is $Q_i$" ($i = 1, \cdots, m$) that relate variables pertaining to the attribute values of some item $x$, and where the $P_i^j$'s and $Q_i$ are classical subsets in the corresponding attribute domains. Then, it has been shown that the result of their joint application (including the fusion of the results obtained from each rule) can be put under the form of a $\min - \max$ matrix product [7]; see [1] for the general case.

The output of this $\min - \max$ product is a possibility distribution over a collection of mutually exclusive alternatives (induced by weighted conclusions on the $Q_i$'s).

A cascade of such min-max products of matrices has a structural resemblance with a min-max neural network. In fact such a cascade can be shown to be equivalent to a min-max neural net, each matrix product corresponding to a layer, and the activation function used being the identity; [1] for details. First, this indicates that there is not such a big gap between experts systems and neural nets, in spite of what some people say or write. Second, this opens the door to the learning of the matrices corresponding to each layer; clearly a gradient descent-like algorithm is not easy here due to the non differentiable nature of the min-max products; however another road may use fuzzy relation equation solving techniques [4].

Besides, the above matrix calculus was recognized early as being of interest for explainability purposes [8]; this early work has been recently pursued [2, 3]. Indeed a sensitivity analysis is made possible by exploiting the expression of the min-max matrices product. Depending on what is fixed and what is known, various forms of explanations can take place (including 'why' and 'why not' queries). See [9] for a detailed example with a cascade of two layers of rules. Generally speaking, the logical side of possibility theory has a potential for explanatory capabilities.

## 3 Analogical setting

Rather than dealing with *rules*, as in the previous approach, one may deal with a repertory of *cases*. A convenient way of expressing analogical inference [5] is to build analogical proportions, which are statements of the form "$a$ is to $b$ as $c$ is to $d$", which is denoted $a : b :: c : d$. Analogical proportions can be encoded as logical expressions stating that $a$ differs from $b$ as $c$ differs from $d$ and that $b$ differs from $a$ as $d$ differs from $c$, see, e.g. [10]. The analogical proportion $a : b :: c : d$ is true only for the 6 following patterns (in the Boolean case): 0 : 0 :: 0 : 0 ; 1 : 1 :: 1 : 1 ; 0 : 1 :: 0 : 1 ; 1 : 0 :: 1 : 0 ; 0 : 0 :: 1 : 1 and 1 : 1 :: 0 : 0 . This readily extends to nominal values with 3 patterns of the form $s : s :: s : s$, $s : t :: s : t$, $s : s :: t : t$ (where $s$ and $t$ are values of the considered nominal attribute).

The table below exhibits two pairs $(a, b)$ and $(c, d)$ such as $a : b :: c : d$ holds for each attribute value $\mathcal{A}_i$ and for the values of the class $\mathcal{C}$) associated to each of the 4 items $a, b, c, d$. Observe that the attributes can be partitioned in three categories: i) from $\mathcal{A}_1$ to $\mathcal{A}_{j-1}$, $a, b, c, d$ are identical ; ii) from $\mathcal{A}_j$ to $\mathrm{A}_{r-1}$, $a, b$ are identical, as well as $c, d$, but not in the same way ; iii) from from $\mathcal{A}_r$ to from $\mathcal{A}_n$, $a$ differs from $b$ in the same way as $c$ differs from $d$. Such a table can be read in a way oriented towards explanation: The *change* of value of $\mathcal{C}$ from $s$ to $t$ between $a$ and $b$ and between $c$ and $d$ can only be explained, given this table, by the *change* of values of attributes from $\mathcal{A}_r$ *to* $\mathcal{A}_n$.

| | $\mathcal{A}_1 \dots \mathcal{A}_{i-1}$ | $\mathcal{A}_i \dots \mathcal{A}_{j-1}$ | $\mathcal{A}_j \dots \mathcal{A}_{k-1}$ | $\mathcal{A}_k \dots \mathcal{A}_{r-1}$ | $\mathcal{A}_r \dots \mathcal{A}_{s-1}$ | $\mathcal{A}_s \dots \mathcal{A}_n$ | $\mathcal{C}$ |
|---|---|---|---|---|---|---|---|
| $a$ | 1 | 0 | 1 | 0 | 1 | 0 | $s$ |
| $b$ | 1 | 0 | 1 | 0 | 0 | 1 | $t$ |
| $c$ | 1 | 0 | 0 | 1 | 1 | 0 | $s$ |
| $d$ | 1 | 0 | 0 | 1 | 0 | 1 | $t$ |

This change is the same for the pair $(a, b)$ and the pair $(c, d)$). So these pairs may be viewed as instances of a rule expressing that the change on attributes from $\mathcal{A}_r$ to $\mathcal{A}_n$ determines the change for $\mathcal{C}$ whatever the context; obviously this rule may have exceptions in the case repertory! However, the building block for finding explanations is the comparison of cases. In fact, it can be shown that one can approximate both abductive explanations (how can we insure that the class of item $x$ is $s$?) and contrastive explanations (why item $x$ is not in class $s$: one can then identify the attributes whose value should be changed for that), in this setting [11]. Besides, the comparison of items can be also at the basis of an index for estimating the relevance of attributes in a classification process [12].

## 4 Conclusion

We have only outlined the main features of two approaches to explanation. The reader is referred to the cited papers for details and examples.

## References


1. I. Baaj, J. Poli, W. Ouerdane, N. Maudet. Min-max inference for possibilistic rule- based system. Proc. 30th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE21), Luxembourg, July 11-14, . 1-6. IEEE. 2021.
2. I. Baaj, J. Poli, W. Ouerdane, N. Maudet. Representation of explanations of pos- sibilistic inference decisions. Proc. 16th Europ. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU21), (J. Vejnarov et N. Wilson, eds.), Prague, Sept.21-24, LNCS 12897, 513-527. Springer, 2021.
3. I. Baaj. Explainability of possibilistic and fuzzy rule-based systems. PhD thesis, Sorbonne University, Paris, 2022.
4. I.Baaj. Max-min learning of approximate weight matrices from fuzzy data. CoRR abs/2301.06141, 2023.
5. M. Bounhas, H. Prade, G. Richard. Analogy-based classifiers for nominal or numerical data. Int. J. Approx. Reason. 91, 3655, 2017.
6. L. Charnay, J. Dibie, and S. Loiseau. Validation and explanation. In P. Marquis, O. Papini, and H. Prade, editors, A Guided Tour of Artificial Intelligence Research. Vol. 1: Knowledge Representation, Reasoning and Learning, (P. Marquis, O. Papini, H. Prade, eds.), Springer, 707-731, 2020.
7. D. Dubois, H. Prade. From possibilistic rule-based systems to machine learning - A discussion paper. Proc. 14th Int. Conf. on Scalable Uncertainty Management (SUM20), (J. Davis, K. Tabia, eds.), Bozen-Bolzano, Sept.23-25, LNCS 12322, 3551. Springer, 2020.
8. H. Farreny, H. Prade. Positive and negative explanations of uncertain reasoning in the framework of possibility theory. Proc. 5th Conf. on Uncertainty in Artificial Intelligence (UAI89), Windsor, ON, Aug. 18-20, 95101, 1989. Available in CoRR, abs/1304.1502, 2013. Revised version in: Fuzzy Logic for the Management of Uncertainty (L. A. Zadeh, J. Kacprzyk, eds.), Wiley, 319-333, 1992.
9. H. Farreny, H. Prade. Explications de raisonnements dans lincertain. Revue dIntelligence Artificielle, 4 (2), 4375, 1990.
10. H. Prade, G. Richard. Homogenous and heterogeneous logical proportions. IfCoLog J. of Logics and their Applications, 1(1): 1–51, 2014

11. S. Lim, H. Prade, G. Richard. Using analogical proportions for explanations. Proc. 15th Int. Conf. on Scalable Uncertainty Management (SUM'22), (F. Dupin de Saint-Cyr, M. Öztürk-Escoffier, N. Potyka, eds.), Paris, Oct. 17-19, LNCS 13562, Springer, 309-325, 2022.
12. S. Lim, H. Prade, G. Richard. Analogical relevance index. CoRR abs/2301.04134, 2023.
13. T. M. Strat. The generation of explanations within evidential reasoning systems. Proc. 10th Int. Joint Conf. on Artifical Intelligence (IJCAI'87), Milano, 1097-1104, 1987.