



HAL
open science

Contributions to Explainable Artificial Intelligence: Developing ExpressIF® for Industry and Sciences

Jean-Philippe Poli

► **To cite this version:**

Jean-Philippe Poli. Contributions to Explainable Artificial Intelligence: Developing ExpressIF® for Industry and Sciences. Artificial Intelligence [cs.AI]. Université Paris-Saclay, 2023. tel-04615709

HAL Id: tel-04615709

<https://cea.hal.science/tel-04615709v1>

Submitted on 18 Jun 2024

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.

Contributions to Explainable Artificial Intelligence: Developing ExpressIF[®] for Industry and Sciences.

*Contributions à l'Intelligence Artificielle Explicable : Développement
d'ExpressIF[®] au Service de l'Industrie et des Sciences.*

Habilitation à Diriger des Recherches

Présentée et soutenue à Saclay,
le 1er Décembre 2023, par

Jean-Philippe Poli

Composition du jury

Anne Laurent Professeure Université de Montpellier, ISDM	Rapporteur
Marie-Jeanne Lesot Professeure Sorbonne Université, LIP6	Rapporteur
Sébastien Destercke Directeur de recherche CNRS, Université de Technologie de Compiègne, HEUDIASYC	Rapporteur
Salem Benferhat Professeur Université d'Artois, CRIL	Examineur
Nicolas Maudet Professeur Sorbonne Université, LIP6	Examineur
Vincent Mousseau Professeur CentraleSupélec, Université Paris-Saclay, MICS	Examineur
Céline Hudelot Professeure CentraleSupélec, Université Paris-Saclay, MICS	Mentor

Contents

1	Introduction	5
1.1	Motivations	5
1.2	Contributions	9
1.3	Context	15
1.4	Structure and content of the document	15
2	Background on Fuzzy Systems	17
2.1	Fuzzy sets	17
2.2	Fuzzy relations	24
2.3	Knowledge representation with fuzzy logic	27
2.4	Reasoning in fuzzy logic	30
2.5	Fuzzy rule-based systems	33
2.6	Fuzzy Constraint Satisfaction Problems	36
3	Improving expressiveness of fuzzy systems	41
3.1	Expressiveness of AI systems	42
3.2	Online temporal predicates	43
3.3	Spatial relations	54
3.4	Online spatio-temporal predicates	58
3.5	Knowledge about the predicates	64
3.6	Applications	66
3.7	Summary	69
4	Extracting knowledge from data for building fuzzy systems	71
4.1	Motivations	72
4.2	Interpretable feature construction	74
4.3	Relevant predicate learning	85
4.4	Extraction of gradual rules from experimental data	92
4.5	Summary	99
5	Improving usability and acceptance of fuzzy systems	101
5.1	Touch graphical user interface for rule authoring	101
5.2	Generation of textual explanations of fuzzy systems decisions	109
5.3	Summary	120
6	Bringing fuzzy systems closer to digital instrumentation	123
6.1	Fuzzy logic for digital instrumentation	124
6.2	Using knowledge to build interpretable features	125
6.3	Adapting fuzzy decision trees to specific data	130
6.4	Summary	138
7	Conclusion and perspectives	143

Conclusion	143
Appendices	161
A Short resume	165
B Teaching activities	167
C Collective responsibilities	171
D Supervision of young researchers	173
E Research projects and fundings	179
F Complete list of publications	181
G Selection of articles	189

List of Figures

1.1	Partial taxonomy of the types of decision an AI can achieve.	6
1.2	Overview of expert systems.	8
1.3	Trade-off between model interpretability and performance.	9
1.4	Contributions to expert systems field.	10
1.5	Example of evaluation graph for a simple fuzzy rule base.	11
1.6	Modular architecture overview (Poli and Boudet, 2018).	11
1.7	Research activities timeline.	16
2.1	The different characteristics of a fuzzy set X	20
2.2	Examples of application of the main t-norms on two fuzzy sets.	21
2.3	Examples of application of the main t-conorms on two fuzzy sets.	22
2.4	Example of the negation (complement) of a fuzzy set.	23
2.5	Example of fuzzy dilation. Unusually, black stands for 0 and white for 1.	27
2.6	Example of a linguistic variable "Temperature"	29
2.7	Plot of the truth values of the core implications according to $\mu_A(x)$ and $\mu_B(y)$	32
2.8	Different applications of GMP for $\top = \min$ and \mathcal{I}_{BG}	34
2.9	Illustration of the core defuzzification operators applied on a fuzzy set.	35
3.1	Membership functions representing two different temporal scopes.	44
3.2	Input signal of a temperature sensor, a membership function associated to its domain, and evaluation of the resulting fuzzy proposition.	45
3.3	The fuzzy proposition "the rainfall is high".	45
3.4	Application of the <i>Occ</i> operator to the expression "the temperature is high" on two different scopes.	46
3.5	Application of <i>Ratio</i> operator to the expression "the temperature is high" on two different scopes.	47
3.6	Principle of using the gradient to characterize growth or decline.	48
3.7	Examples of growth modifiers and their interpretations.	49
3.8	Examples of decline modifiers and their interpretations.	50
3.9	Application of the <i>Pers</i> operator to the expression "the temperature is high" on two different scopes.	51
3.10	Application of the <i>PersWhile</i> predicate to the expression "the temperature is high" with a persistence of 5 seconds during the last 20 seconds.	52
3.11	Application of the <i>PersWhileTol</i> predicate to the expression "the temperature is high" with a ratio computed on 5 seconds during the last 20 seconds.	52
3.12	Evaluation of the expression "the rainfall is high has started before that the temperature is high" on two different scopes.	53
3.13	Evaluation of the expression "the rainfall is high before the temperature is high" on two different scopes.	53
3.14	Two structuring elements ν_G^P representing the relation "in the direction G".	55
3.15	Fuzzy landscape of two objects (crisp in this example) dilated by ν_G	56
3.16	Left: membership function for the "Near" relation. Right: corresponding structuring element ν_n with $\delta_1 = 4$ and $\delta_2 = 9$	57

3.17	The same objects as on Figure 3.15 dilated by ν_{NG}	57
3.18	Membership function "at ground level" with respect of the computed value of Δ_z for any point X of \mathbb{R}^3	58
3.19	An excerpt of the DTM and the ground level.	58
3.20	Example of a membership function used for estimating "the distance traveled by e since the previous position is not null".	59
3.21	Elements of comprehension for the <i>IsGoingCloseTo</i> predicate.	60
3.22	Illustration of the different elements of the <i>IsFollowingAPath</i> operator equation.	63
3.23	Different examples of trajectories which cross the region	63
3.24	Illustration of the different elements of the <i>IsCrossing</i> operator equation.	64
3.25	Subsumption lattice for the spatio-temporal relations	65
3.26	Graph representing the logical links between the nine spatial predicates considered in the example	66
3.27	Representations of the runoff area.	67
3.28	Examples of runoff from crisp polygons, with various shapes and from different points on the DTM.	68
3.29	View of the Corte castle (from South).	68
3.30	Runoff area from the Corte castle.	68
4.1	Inside of an atom	74
4.2	Overview of the principle of genetic programming	76
4.3	Grammar used for the CLAS12 dataset.	77
4.4	Results of the survey on constrained against unconstrained feature construction.	82
4.5	Results of the survey on prior against embedded feature construction (FC).	82
4.6	Results of the survey on automatic feature construction (FC) against regular variables used by physicists.	83
4.7	The nine organs of interest in this experiment (in color).	91
4.8	Overview of materials science	92
4.9	Overview of the proposed approach for gradual rule extraction.	93
4.10	GGMP decomposition of a triangular fuzzy set A	95
4.11	Results on a toy dataset.	97
5.1	Examples of GUI to edit rules.	102
5.2	Overview of ExpressIF [®] <i>Rule editor</i>	103
5.3	The choice of shapes for a membership function.	103
5.4	Example of a linguistic variable and its terms. The points are the handles to move.	104
5.5	Focus on the possible destinations while <i>drag-and-drop</i> of the binary operator.	104
5.6	Pie menu for binary operator selection.	105
5.7	Duration (in s) of exercise 1 per education level (undergraduate or graduate).	106
5.8	Error rate of all users, per exercise and per tool (in %).	108
5.9	Overview of the proposition of textual explanation generation.	110
5.10	Backward MRI image with different regions to annotate.	116
5.11	Results of the evaluation of the explanations generated in the pasta case study.	119
5.12	Results of the evaluation of the explanation generated in the MRI semantic annotation case study.	120
6.1	Principle of Rayleigh-wave SAW sensors.	126
6.2	Examples of frequency shifts due to exposition to different compounds.	126
6.3	Concentration setpoint and concentration profile estimate.	129

6.4	Screenshot of ClooNEZ software.	129
6.5	Overview of tagged neutron system in C-BORD project	131
6.6	Impact of the degree of uncertainty on the accuracy on a toy dataset, for uncertain fuzzy decision trees and regular fuzzy decision trees.	134
6.7	Example of a fuzzy decision tree for material classification.	135
6.8	Confusion matrix from a cross-validation on balanced Saclay dataset for the (matrix, target) classification.	136
6.9	Confusion matrix from a cross-validation on Rotterdam data for the four-class and mixtures problem.	137
6.10	Confusion matrix from a cross-validation on Rotterdam data for the organic three-class problem.	137
7.1	XAI emphasis	144
7.2	Overview of the proposed approach.	145

List of Tables

2.1	Examples of dual t-norms and t-conorms.	23
2.2	Core fuzzy implications in a rule like “if V is A then W is B”. The two last implications do not generalize the classical implication.	31
2.3	Compatibility between the core GMP operators and the fuzzy implications.	33
3.1	The different kinds of link between relations and their notation in the graph representation.	65
4.1	Results of FURIA, obtained with a 5-fold cross-validation on several famous datasets.	73
4.2	Transition matrix for the CLAS12 dataset.	78
4.3	Cohen’s kappa score for one built feature.	79
4.4	Cohen’s kappa score for built features.	80
4.5	Statistics and contamination percentage in three bins with several models.	84
4.6	Evaluation of the prediction of <i>DeS</i> values based on <i>Pr</i> and <i>P</i> , and of <i>Th</i> values based on <i>P</i> , <i>NP</i> and <i>SS</i> , using the different selected predictors.	98
5.1	Average durations for exercise 2 and exercise 3 per education level.	107
5.2	User preferences for a tool per skills.	108
5.3	IPCC uncertainty scale	112
5.4	The pasta classifier rulebase.	113
5.5	Survey questionnaire to evaluate textual explanations.	118
6.1	Typology of problems and approaches for mixture recognition.	128
6.2	Accuracies (in %) of 3 classifiers used either with the amplitudes or our features.	129
E.1	European projects.	180
E.2	National projects.	180
E.3	CEA projects.	180
F.1	Publications per categories.	181

Introductory chapters

[...] Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.

J. McCarthy
M. Minsky
N. Rochester
C. Shannon

Chapter 1

Introduction

This document aims at presenting a summary of my research and the main results I obtained since my doctorate in computer science, defended at Université Paul Cézanne in 2007. This research has been carried out since I arrived at CEA List, an institute of CEA Tech, the same year.

My research consists in providing end-users with an efficient decision tool. In the beginning, it was restrained to approaches when no data is available. I thus focused on expert systems and building models from human expertise. In the concern of allowing the end-users to manipulate directly this tool, my work has been directed towards usability and knowledge representation. The fuzzy logic framework has been chosen for its capacity to handle language vagueness, using words instead of mathematical operators. Later, I interested in building such models directly from data and in providing textual explanations for decisions.

My research is thus at the intersection of different domains of Artificial Intelligence: symbolic artificial intelligence (knowledge representation, reasoning), fuzzy logic, XAI¹, machine learning and ergonomics. The tool we are developing is thus definitely user-centric: it is an XAI that must be usable by end-users.

This work has been applied to practical real-world problems from companies that partly funded it: large companies (e.g. Airbus, Veolia, Orano, Total, etc.), medium companies (e.g. Teréga, InGroupe, etc.) and small companies (e.g. Alicante, MHCComm, etc.). I chose to focus on the methodological aspects since most of the applications are confidential.

In the next section, I will motivate the work presented in this manuscript.

1.1 Motivations

Today's world imposes a competitive impulse that often leads to a need for lower operating costs. In particular, companies, hospitals and laboratories must make decisions faster based on multiple criteria. These decisions must be made very quickly despite increasing complexity:

- A *growing number of criteria*. In particular, Miller estimated the limitation of human memory to 7 ± 2 concepts at the same time (Miller, 1956): this limits the capacity of human experts to make a decision based on a bigger amount of data. Moreover, sometimes, the data come as a flow, which makes the decision process even more difficult.
- The *heterogeneity* of the criteria, which has a double meaning. First, it can be applied to the presence of different natures of information (e.g. images, sounds, scalars, etc.) during the decision process. Then, it is also used for group decision making when the

¹eXplainable Artificial Intelligence, (Gunning and Aha, 2019)

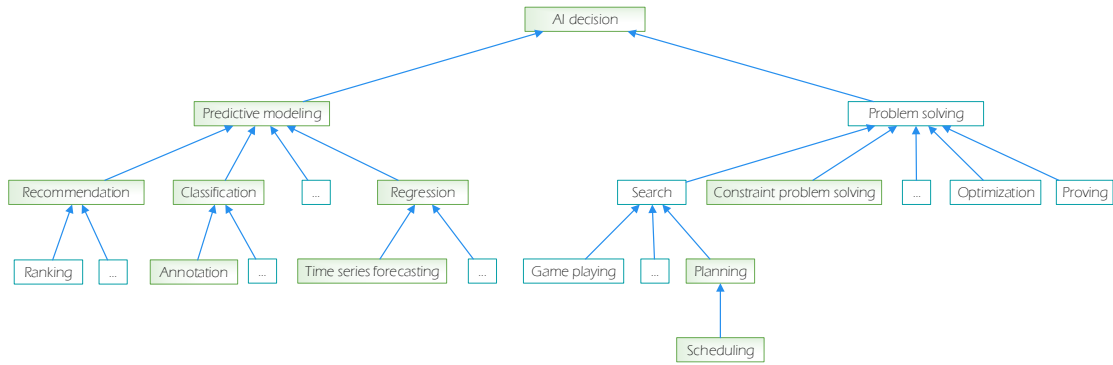


Figure 1.1: Partial taxonomy of the types of decision an AI can achieve. The colored tasks are those we have worked on.

profiles of decision makers are different (e.g. multidisciplinary consultation meetings in hospitals).

- The *combinatorial* complexity. Humans have shown poor performances in solving combinatorial problems, compared to Artificial Intelligence, especially when the problem is large. This relates to the great successes of the "good old-fashioned AI" with constraint-solving problems, planning, search, etc.
- The *imperfection* of data, knowledge and beliefs involved in the decision process. Unfortunately, in most cases, the decisions are made under uncertainties, which can lead them away from rational decisions.

Under certain conditions, which we will explore later, Artificial Intelligence (AI) can provide good decision tools. In particular, AI offers unbiased and consistent decisions (e.g., no feeling comes to disturb the process) and obviously surpasses the availability of human experts. In general, AI reduces human errors. Indeed, an AI model that is trained on data benefits from a huge amount of different experiences, and a knowledge-based model has all the knowledge at all times (what humans can hardly do, or else with a concentration that cannot be maintained for too long). Thus, AI can naturally deal with the four points presented above:

- most of the AI approaches can deal with a great number of inputs;
- inputs heterogeneity is still a challenge in some fields of AI (e.g. Deep Learning, even if the problem is already tackled) but is undertaken by others (e.g. expert systems);
- AI can solve combinatorial problems more easily than humans do;
- different formalisms have been used to consider imperfections in data and knowledge.

We can distinguish different decision tasks that can be achieved by an AI. Figure 1.1 shows a non-exhaustive taxonomy of such tasks.

On the one hand, predictive modeling consists in finding an approximation \hat{f} of an actual function f and which is a mapping from input variables $x \in \mathcal{X}$ to an output variable $y \in \mathcal{Y}$. For now, we do not consider what kind of mathematical model \hat{f} is, and how its parameters, if any, have been set. Regarding the nature of \mathcal{Y} , we can distinguish classification (when \mathcal{Y} is discrete) that consists in the assignment of a label to the data, and regression (when \mathcal{Y} is continuous) that rather computes a quantity. In the latter case, time series forecasting is a special case where data are ordered by time. We also find recommendation that consists in predicting the next item that can interest a user.

On the other hand, AI can solve problems and the decisions have as many forms as there are types of problems. Solving is a general type of approach that consists in finding a sequence of actions to reach a goal. Planning, scheduling and game playing can be viewed as particular cases of search. The solving of constraint satisfaction problems leads to the assignment of values to variables that satisfy the constraints, if at least one exists, or to a failure otherwise (see section 2.6).

Moreover, AI can perform these tasks with different levels of autonomy:

- In *decision support*, the AI extracts insights from the current situation that give clues for a human expert to make the decision. The AI is barely involved.
- In *decision augmentation*, the AI provides the human decision maker with a fewer number of alternatives. Roughly, both the human expert and the AI participate equally to the decision making.
- Finally, in *decision automation*, the AI makes the decision alone.

Based on these observations, we propose augmented decision-making to companies, hospitals and laboratories by relying on what they have: data and/or expertise. In particular, we are not targeting only big data because some partners cannot pretend to collect such an amount of data (e.g. businesses that have just started, experimental data that are too long or too expansive to collect, etc.) and because some of the scenarios are not conducive to big data (e.g. anomaly detection). Since the user is still involved in the decision process, we need to pay a special attention to both *usability* and *acceptance*. On the one hand, usability is related to ergonomics and to how easily a decision maker can use the decision-helping tool. On the other hand, acceptance is more about how decision makers will develop confidence in the tool thank to the trustworthiness of its decisions. For instance, the authors in (Schneider and Xhafa, 2022) point out that some black-box² models used in the medical field led to a negative opinion from the caregivers.

To respond to the problems of acceptability and usability described above, we claim that symbolic AI is a good candidate to decision augmentation. Symbolic AI is the subfield of AI that uses high-level symbols, which are understandable by humans, to define knowledge and reasoning. It can achieve all the tasks that are presented in figure 1.1 and it can address the growing complexity of the decisions.

As a result of our research, we created a framework called ExpressIF[®] that aims at allowing its users to build symbolic AIs for decision augmentation, with regard to usability and acceptance. We chose fuzzy logic as the formalism to represent the knowledge. ExpressIF[®] includes more and more reasoning types: in this manuscript, we focus on Fuzzy Inference Systems (FIS), that are versatile enough to be used with data and/or knowledge, and that can be used for predictive modeling (see chapter 2, section 2.5 for a presentation of FIS). Chapter 4 also considers quickly Fuzzy Constraint Satisfaction Problems (FCSP), which are introduced in chapter 2, section 2.6.

In particular, ExpressIF[®] allows building expert systems whose principles are shown in figure 1.2. In the core of the system, an engine applies the problem-specific knowledge contained in the knowledge base to observations in order to output a decision. The life cycle of an expert system can be split into two main phases: the setup and the consultation (i.e. exploitation). A third phase could be the update of the knowledge base, for instance either manually or by incremental learning.

The first phase consists in gathering all the elements that will be necessary to set up the system. It can be historical data, documentation, knowledge management tools (e.g. wiki) or directly experts. It also consists in deciding what kind of decisions the system

²*Black box* designates AI models that are not interpretable, i.e. which cannot be understood by Humans (Doshi-Velez and Kim, 2017). A proper definition of interpretability will be given in chapter 4.

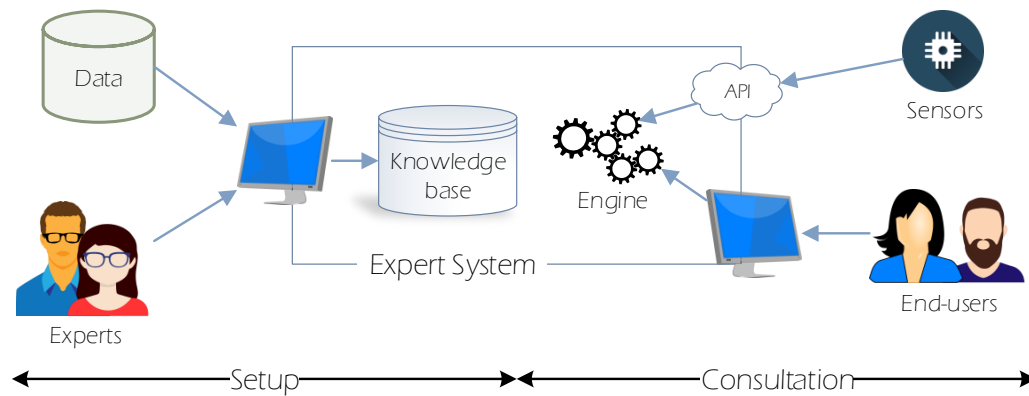


Figure 1.2: Overview of expert systems.

should make. Indeed, it happens quite often that the materials we are provided with do not meet the requirements of the tasks. For instance, in our experience, we have been asked to predict a time to failure for a complex system, only based on expert interviews. Unfortunately, the experts were not able to quantify the time to failure based on the different situations. The solution has been to express a risk of failure instead. Conversely, historical data would have been helpful for such a quantization. In this phase, usability is of paramount importance. It is important that experts can author their own pieces of knowledge, without needing a cognitive engineer. This involves not only ergonomics but also the ability to check the consistency of the knowledge base. In terms of acceptance, it needs also a way to keep the knowledge base interpretable.

The second phase consists in exploiting the expert system. We experienced two main ways: either an end-user or a sensor queries the system. ExpressIF[®] is also able to process both offline and online, i.e. making decision on the fly. The usability here refers to ergonomics but also to the availability of the system like the different interfaces (e.g. web API). To increase the acceptance of the system and its decisions, a real dialog with the end-users can be necessary, providing them with explanations if needed. This connects us to the field of XAI.

Nevertheless, expert systems and symbolic AI suffer from several drawback (Bobrow, Mittal, and Stefik, 1986; Hardaway and Willi, 1990):

- The *knowledge acquisition* has been one of the biggest limit of expert systems. Traditionally, knowledge engineers interview domain experts to formalize their expertise. It happened to be time consuming and costly. Let us quote (Bobrow, Mittal, and Stefik, 1986):

"One of the dreams of the expert-system community is to eventually have knowledge bases created and maintained by their users rather than by knowledge engineers. A dream that reflects certain financial realities."

- *Symbolic learning is still struggling* compared with connectionism. This is partly due to the fact that symbols have been introduced first to be readable by humans, and so not intended to machines (Sun, 2015). However, rule induction and inductive logic programming do exist.
- *The most interpretable models suffer from lower accuracies* (Barredo-Arrieta et al., 2020), as figure 1.3 illustrates. The figure shows also the mistaken reputation of rule-based models that, in the opinion of the authors, are less accurate than decision trees whereas they are intimately linked (Quinlan, 1993).

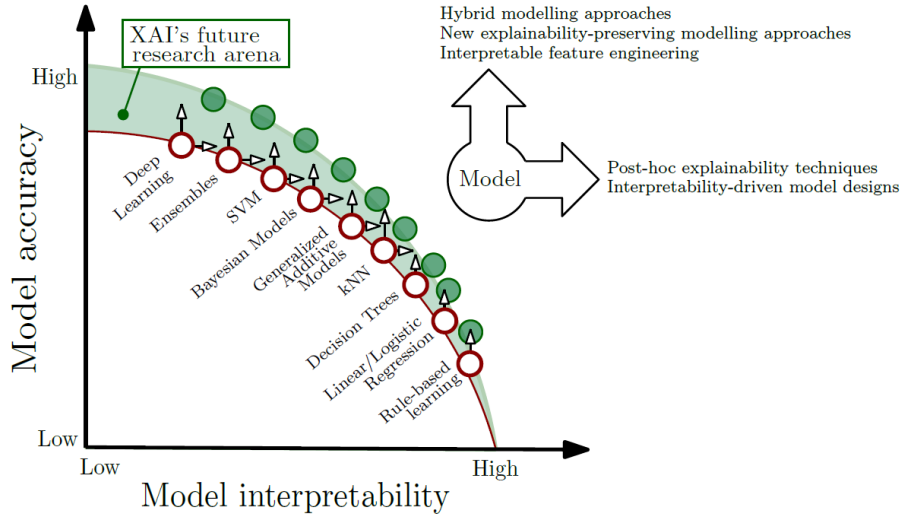


Figure 1.3: Trade-off between model interpretability and performance (Barredo-Arrieta et al., 2020).

- *The acceptance of AI systems*, beyond the ethical issues, is of paramount importance. It depends on many factors, such as the perceived usefulness, the performance expectancy and trust (Kelly, Kaye, and Oviedo-Trespalcios, 2023).

Despite all this, symbolic AI, and more precisely fuzzy systems, meet most of the requirements of interpretability that are described in (Barredo-Arrieta et al., 2020) or in (Lipton, 2018), which is a step of paramount importance for acceptance (Alonso Moral, Castiello, et al., 2021; Bouchon-Meunier, Lesot, and Marsala, 2021).

ExpressIF[®] must be an alternative to the existing frameworks, which helps building trustful augmented decision systems, with the user in-the-loop at each stage, and based on fuzzy logic. In the next section, I will describe my contributions towards such a tool.

1.2 Contributions

Our goal is to provide our partners (e.g. companies, hospitals, laboratories, etc.) with a solution for augmented decisions. Our approach must be multipurpose in the sense that it can be applied to a multitude of scenarios and should solve as many problems as possible. In particular, for the usability and acceptance of this tool, it is of paramount importance it may be customizable and may be able to provide explanations of its decisions. In our case, this tool is ExpressIF[®], which is the implementation of all our research activities.

I started the implementation of ExpressIF[®] with technical contributions (section 1.2.1) that serve as a basis for the others (sections 1.2.2, 1.2.3, 1.2.4). We end this section with contributions in the field of digital instrumentation (section 1.2.5). Figure 1.4 maps our contributions with the expert system schema.

1.2.1 Extensible architecture and rule base representation

Fuzzy systems and fuzzy rule bases will be defined in chapter 2. In our system, the rule base is represented by a directed acyclic graph $G = (V, E)$ where each vertex in V represents a sub-expression and each edge in E represents a dependency (Poli and Boudet, 2018). For instance, let n_1 and n_2 be two vertices, then $n_1 \rightarrow n_2$ means that if the value of n_1 changes, then n_2 must be reassessed (Poli and Boudet, 2016a, 2018). When the rule base is built, we tackle the problem of common sub-expression elimination, building a unique

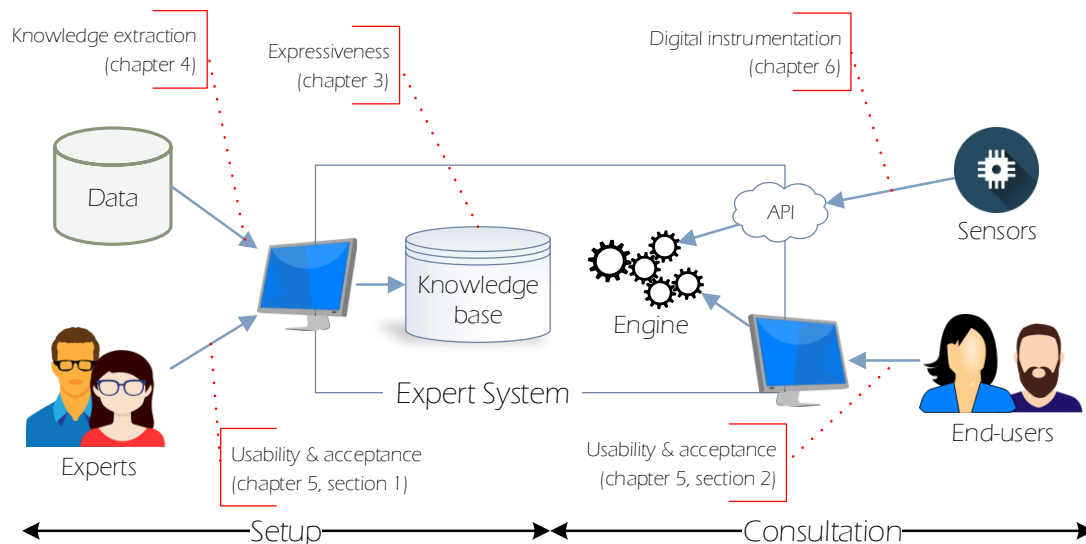


Figure 1.4: Contributions to expert systems field.

node to represent the different instances of a sub-expression. Each node is in charge of its own evaluation and of firing the evaluation of its parents if needed. This allows extending seamlessly the engine: adding a new operator (predicates, relations, etc.) consists in creating a new type of node. The evaluation of the rule base is hence based on an algorithm that propagates the changes from the inputs to the outputs. The propagation stops when a node is not changing: this limits the overall computation time and allows handling input streams. Figure 1.5 shows an example of a small fuzzy rule base represented in ExpressIF[®] as an evaluation graph. The number between the brackets gives the evaluation priority of each node.

We also suggested a new architecture for fuzzy inference systems, as shown in figure 1.6 (Poli, 2016; Poli and Boudet, 2015, 2016a). The goal was to get a system that can process information streams or static data and that can be used in various scenarios. This modularity is ensured by a separation of the tasks and a customization provided by the use of policies. A policy is a set of parameters that customize the behavior of each module. The combination of the behaviors of all the modules enables to address a lot of applications and issues: regular or irregular data rate, delay before inference, etc. The architecture is composed of several modules. The *active input queue* gathers the input and groups them by timestamp. The *scheduler* is able to monitor the system (via the operating system) and to decide which inputs group has to be processed. The *evaluator* is in charge of the evaluation of the rules. Finally, the *output change broadcaster* informs the user about outputs changes and provides eventually an explanation.

The different modules aim at avoiding an overload of the system (for instance, the active input queue selects the inputs which should be treated) or user overfeeding (for instance, the output change broadcaster displays only the relevant information).

Projects	MobiSIC, Edens
Publications	Poli, 2016; Poli and Boudet, 2015, 2016a, 2018

With the modularity of our system, we then worked to improve the expressiveness of fuzzy inference systems.

1.2.2 Expressiveness

This research axis is about using fuzzy logic to model new predicates (see chapter 2, section 2.3) that will be used in the rules and in explanations. The idea behind is to model

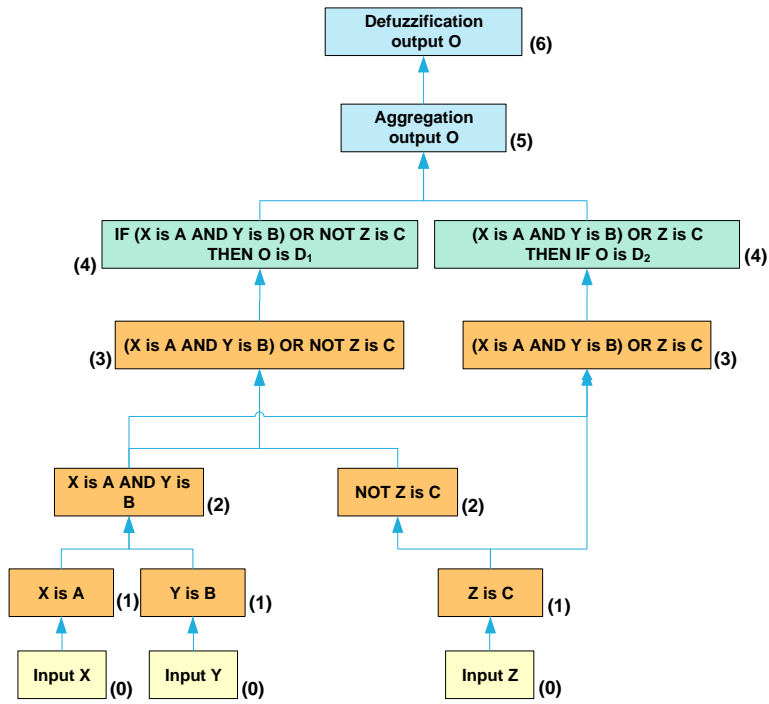


Figure 1.5: Example of evaluation graph for a simple fuzzy rule base.

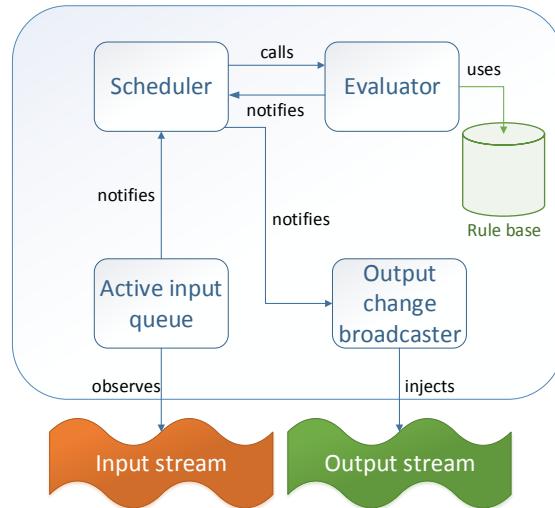


Figure 1.6: Modular architecture overview (Poli and Boudet, 2018).

expert knowledge seamlessly by using such high-level predicates. Each predicate maps a word (or a sentence fragment) with a mathematical formulation of its degree of fulfillment. This axis targets the bottleneck of knowledge formalization. In our opinion, if experts have to transfer their knowledge into an expert system, it must be expressible as close to their thoughts as possible. Moreover, the rewording of this expertise can lead to mistakes.

Improving expressiveness has another virtue: the textual version of the predicates can be used to build a textual explanation of the decision. In particular, a predicate can be added to an explanation if it is involved in one of the knowledge fragments that has been used during reasoning.

During my postdoctoral fellowship, I introduced the mechanism of expiration (see chapter 3, section 3.2.3), which allows considering information streams. Based on this mechanism, I defined several temporal predicates for online causal systems. They are particularly

useful for maintenance and diagnostics applications (Poli, Boudet, Espinosa, et al., 2017). They express notions like persistence, occurrence and variation characterization (Poli and Boudet, 2016b; Poli, Boudet, and Mercier, 2016).

Based on the temporal predicates and spatial predicates, which have been studied by Laurence Boudet for her own projects, I then supervised Jean-Marie Le Yaouanc during his postdoctoral fellowship on spatio-temporal predicates applied to security and fleet management projects (Le Yaouanc and Poli, 2012; Poli, Boudet, and Le Yaouanc, 2018). The goal was to recognize specific behaviors of geolocalized entities relatively to spatial zones.

Finally, I co-supervised the postdoctoral fellowship of Clément Iphar with Laurence Boudet during the Respondrone project, with a crisis management application. His work consists in the creation of spatial predicates in 2.5D in order to be able to reason on digital terrain models. Indeed, the 2D representation of space is not always sufficient to characterize events such as flooding or even wildfires that obviously are dependent from the slopes of the ground (Iphar, Boudet, and Poli, 2021a,b, 2023a).

Supervision	Postdoc: Jean-Marie Le Yaouanc, Clément Iphar Interns: Tristan Galliez, Etienne Pétreil, Julien Plouvier, Sylvain Jankowiak
Projects	Egidium, Airbus, Crédit Agricole, Véolia MobiSIC, Descartes, Edens Respondrone, DeepHealth, Narsis
Publications	Iphar, Boudet, and Poli, 2021a,b, 2023a,b; Le Yaouanc and Poli, 2012; Pierrard, Cabaret, et al., 2020; Poli and Boudet, 2016b; Poli, Boudet, Espinosa, et al., 2017; Poli, Boudet, and Le Yaouanc, 2018; Poli, Boudet, and Mercier, 2016

1.2.3 Knowledge extraction

With the proliferation of datasets, a modern AI tool should be able to train from data. In human-in-the-loop scenarios we aim at, we develop novel algorithms that extract knowledge from data, which must meet the two following criteria. First, they must provide valuable insights to users about the data by using fuzzy relations and interpretable rules and constraints. Then, they must be executable by an inference engine or a solver.

That is why we talk about knowledge extraction instead of model induction. We carried out research to address two main kinds of problems.

The first one started from the observation that most of the interpretable models were not able to build their own representation of the problem. We know that this is one of the strengths of Deep Neural Networks. We thus suggested building automatically higher-level features, based on the dataset raw features, under the constraint of interpretability (Cherrier, Poli, Defurne, and Sabatié, 2019). In particular, we forced them to respect mathematics, their units and their dimensions. For instance, our method would not build a new feature by adding a weight and a size. Feature construction is performed either a priori, i.e. as a way to augment the dataset columns, or in an embedded way, during the induction of the model itself (Cherrier, Defurne, et al., 2019). The embedment is specific to each model and we adapted decision trees (crisp and fuzzy), FURIA (Cherrier, Poli, Defurne, and Sabatié, 2020) and Generalized Additive Models (Cherrier, Mayo, et al., 2020). This was the goal of Noëlie Cherrier’s PhD, in collaboration with Maxime Defurne and Franck Sabatié from CEA Irfu, applied to High Energy Physics, when we also surveyed the interpretability of both features and models.

The second aspect of my research in this domain is to be able to learn rules whose semantics is higher than conjunctive ones. During Régis Pierrard’s PhD, directed by Céline Hudelot (MICS, CentraleSupélec), we explored rules based on fuzzy predicates. The fuzzy predicates were selected automatically (Pierrard, Poli, and Hudelot, 2018a) to extract

either rules, for image classification, or constraints, for image object annotation (Pierrard, Poli, and Hudelot, 2018b; Pierrard, Poli, and Hudelot, 2019). The use of predicates in automatically created rules is also helping for the explanation of decisions made by such rule bases (Pierrard, Poli, and Hudelot, 2021). With the postdoctoral fellowship of Hiba Hajri, followed by the internship of Killian Susini, we aimed at extracting gradual rules³ from experimental data (Hajri, Poli, and Boudet, 2021b,b). Indeed, they carry a very strong semantics, in particular in the domain of material science. Graduality has different meanings and representations in fuzzy logic and we tested several ones in order to build a method that works on few or big data, eventually with noise. This work was the first occasion to use causality as a way to select inputs related to the outputs to build stronger rules that match better with the way humans learn and reason.

More recently, I continued my efforts to study different aspects of causality (Jacquin, Lomet, and Poli, 2021; Kunitomo-Jacquin, Lomet, and Poli, 2022). This work is achieved with Lucie Kunitomo-Jacquin and Antonin Arsac, both co-supervised with Aurore Lomet. Our goal is to find an efficient way to extract causality from data or time series, with as less as possible hypotheses about the data.

This axis is thus related to all the bottlenecks that have been mentioned previously.

Supervision	PhD: Régis Pierrard, Noémie Cherrier, Antonin Arsac Postdoc: Hiba Hajri, Lucie Kunitomo-Jacquin, Nadia Ben Abdallah, Olivier Rousselle Engineer: Arnaud Grivet Sébert Interns: Théo Rubenach, Thomas Lamson, Killian Susini Apprentices: Lucas Payet, Kevin Gallus
Projects	Total Energies CBORD, DeepHealth, Entrance Essaim, Alcryphe, providIA, DIAMANT, C3PO
Publications	Arsac, Lomet, and Poli, 2023; Cherrier, Defurne, et al., 2019; Cherrier, Mayo, et al., 2020; Cherrier, Poli, Defurne, and Sabatié, 2019, 2020; Grivet Sébert and Poli, 2018a; Hajri, Poli, and Boudet, 2021b,b; Jacquin, Lomet, and Poli, 2021; Kunitomo-Jacquin, Lomet, and Poli, 2022; Pierrard, Poli, and Hudelot, 2018a,b, 2021; Pierrard, Poli, and Hudelot, 2019; Rousselle, Poli, and Ben Abdallah, 2023

1.2.4 Usability and acceptance

This research axis concerns the usability of a fuzzy inference system by end-users, and so the first and the last bottlenecks. It is strongly related to the industrial transfers.

To improve the usability of such systems, I first worked on ergonomics, in particular for rule authoring, during Jean-Paul Laurent’s internship. We used modern interfaces, like touch screens, to develop a rule editor based on drag-and-drop to decrease the cognitive load during authoring (Poli and Laurent, 2015a,b). The interface guides the end-users in order to obtain a well-formed rule base (Poli and Laurent, 2016). Nevertheless, this is not the only functionality an end-user needs. We worked on rule base validation during the internship of Martin Everaert, co-supervised with Edwin Friedmann. Interactively, the rule base validator provides warnings and errors as interactive compilers do in integrated development environments.

To raise fuzzy systems to the level of XAI, I have worked on textual explanation generation in order to increase the trust in the decisions during the exploitation phase (Baaaj and Poli, 2019). We investigated this problem during Régis Pierrard’s PhD (Poli, Ouerdane, and Pierrard, 2021a,a), and more specifically during Ismaïl Baaaj’s PhD, within a collaboration with Wassila Ouerdane (MICS, CentraleSupélec) and Nicolas Maudet (LIP6, Sorbonne Université) (Baaaj, Poli, Ouerdane, and Maudet, 2021a,b; Baaaj, Poli, and Ouerdane, 2019).

³A gradual rule has the form “the more ... the more...” or any combination of less and more.

The novelty of our approach resides in the use of natural language generation to avoid templated explanations. We also introduced a questionnaire to evaluate the quality of an explanation (BaaJ and Poli, 2019).

Supervision	PhD: Ismaïl BaaJ Interns: Jean-Paul Laurent, Martin Everaert, Ismaïl BaaJ Apprentices: Stéphane Barat, Bastien Guillon, Sébastien Klasa
Projects	Alicante Edens, BIMEET, BIM4VET, ResponDRONE, DeepHealth, Narsis
Publications	Aupetit et al., 2015; BaaJ, Poli, Ouerdane, and Maudet, 2021a,b; BaaJ and Poli, 2019; BaaJ, Poli, and Ouerdane, 2019; Poli, 2016; Poli and Boudet, 2015, 2016a, 2018; Poli, Boudet, Espinosa, et al., 2019; Poli and Laurent, 2015a,b, 2016; Poli, Ouerdane, and Pierrard, 2021a,a

1.2.5 Digital instrumentation

Finally, I have worked with a service at CEA List that develops sensors (mainly nuclear and chemical sensors). They usually used classification methods that are related to kNN (and sometimes with $k = 1$). I usually provide controllers for their sensors and apply machine learning methods to leverage their sensors as intelligent sensors (Boudergui et al., 2011; Magne et al., 2021; Pino, Fontana, Nebbia, Pedersen, et al., 2021). In particular, during the PhD of Olivier Hotel, we collaborated with Samuel Saada, Christine Mer and Emmanuel Scorsone from CEA List to develop new algorithms to tackle different problems with chemical sensors (Friedmann and Poli, 2019; Friedmann, Poli, et al., 2020; Hotel, Poli, Mer-Calfati, et al., 2017a,a,c, 2018; Hotel, Poli, and Saada, 2017).

I supervised different persons to work on the processing of digital instrument data: Rabah Abdul Khalek (intern) and Marouen Baalouch (postdoctorate) in collaboration with Franck Sabatié and Maxime Defurne (Baalouch et al., 2019). We investigated with them the use of neural networks to process High Energy Physics datasets, in order to compare them with the interpretable models developed by Noëlie Cherrier.

During the collaboration with Arnaud Grivet Sébert (non-permanent research engineer), we adapted fuzzy decision tree induction algorithms to consider imperfect data (Grivet Sébert and Poli, 2017, 2018a). The data were modeled as Gaussian distributions by physicists.

This axis is a way to adapt ExpressIF[®] to sensors and to extract symbols from raw data (e.g. peaks from spectra, segments from multi-valued signals). This ensures the continuum from sensors raw data to high-level decisions.

Supervision	PhD: Olivier Hotel Postdoc: Marouen Baalouch Interns: Rabah Abdul Khalek Apprentices: Jérémy Babouche, Florent Brouca, Robin Delgado, Mikaël Gendreau, Sébastien Klasa, Lucas Payet, Kevin Gallus
Projects	Orano, FramAtom Essaim Secured, Scintilla, Safewater, CBORD, Terrific, Spadassin, MI-CADO, Entrance, DetecTool
Publications	Baalouch et al., 2019; Boudergui et al., 2011; Friedmann and Poli, 2019; Friedmann, Poli, et al., 2020; Grivet Sébert and Poli, 2017, 2018a; Hotel, Poli, Mer-Calfati, et al., 2017a,a,c, 2018; Hotel, Poli, and Saada, 2017; Magne et al., 2021; Pino, Fontana, Nebbia, Carasco, et al., 2022; Pino, Fontana, Nebbia, Pedersen, et al., 2021

All the research presented in this section have been achieved under a specific constraints. In the next section, I will describe the specific context of CEA Tech and how I proposed a model that is adapted to applied research and that is suitable in such an environment.

1.3 Context

CEA Tech is the technology division of CEA (the French Atomic Energy and Alternative Energy Commission). It is dedicated to bringing manufacturers a broad range of key enabling technologies and is poised to drive and coordinate the French government’s innovation-led industrial regeneration policy. The mission of CEA Tech is to operate at Technology Readiness Levels (TRLs) 3 to 7, i.e. from the proof of concept to prototype.

CEA Tech’s model has an impact on the way we carry out research: in particular, it is necessary to fund our activities. This means the implementation of a scientific roadmap depends on funding initiatives and opportunities. Consequently, one may interrupt an activity for a while, waiting for a project to be accepted to fund it.

After I arrived at CEA List, I suggested a model to maintain a balance between research, funding and transfer to companies. This model consists in developing a software that is now called ExpressIF[®]. This is particularly necessary for Symbolic Artificial Intelligence, since there is still no major framework compared to Deep Learning with TensorFlow, Keras, pyTorch, etc. Technological improvements and research results combine to update ExpressIF[®] that is transferred, under license, to our industrial partners.

I thus created an activity during my post-doctoral fellowship and I am now in charge of the evolution of the software and the research activities. The ExpressIF[®] team grew over time and we are now six permanent researchers working full-time on it. Our activity consists in growing and maintaining the main multi-purpose platform and applying the existing algorithms to specific domains.

1.4 Structure and content of the document

This manuscript is divided into three parts. The first part consists in the introductory chapters, constituted by this introduction and chapter 2 about the background. Readers that are familiar with fuzzy logic and inference may skip chapter 2.

The second part relates my research activities and describes my contributions. It consists in four chapters that follow the last four axes presented before. Indeed, I have chosen not to describe the architecture and its benefits (Poli and Boudet, 2018). Figure 1.4 helps localizing the chapters in the expert system schema.

Chapter 3, Improving expressiveness of fuzzy systems starts with a definition of expressiveness and vocabulary for knowledge-based systems. Basically, expressiveness is related to the words that such a system can handle. It then presents predicates in the temporal domain and the notion of expiration that allows applying ExpressIF[®] on data and information streams. The chapter then presents predicates in the spatial domain, which are used to reason on a digital terrain model. Coupling both space and time, I then define some spatio-temporal predicates that allow characterizing the activity of an entity. The chapter ends by a thought about the links between the various predicates and how it can be used in machine learning approaches.

Chapter 4, Extracting knowledge from data for building fuzzy systems relates the efforts that have been made for automatically extracting knowledge from data. It underlines the importance of extracting knowledge instead of training a model, which emphasizes the importance of interpretability in our work. The specific interest in knowledge extraction is motivated by an empirical analysis of the best rule learning methods. The chapter describes three novel approaches. The first one was born from the observation that inter-

interpretable models do not build their own representation of the variables. We thus propose to build automatically interpretable features from raw features. Another approach consists in taking advantage of the expressiveness of ExpressIF[®] to learn predicate-based models. The method is illustrated on two use cases: image classification and image object annotation. Finally, the chapter ends with the extraction from data of a specific type of rules: the gradual rules. We then illustrate the importance of such rules in material science.

Chapter 5, Improving usability and acceptance of fuzzy systems addresses the problems of usability and acceptance we introduced before. We illustrate by contributions that aim at improving both of the phases of expert systems: the setup and the exploitation. Indeed, the first contribution introduces a modern way to author rules in a fuzzy inference system. It takes advantage from ergonomics insights, like the use of touch interfaces, the importance of cognitive overload, etc. The second contribution consists in providing the end-users with an explanation. We develop several algorithms regarding the type of models and the type of reasoning that are used to solve a problem.

The last chapter of this part is *Chapter 6, Bringing fuzzy systems closer to digital instrumentation*, in which we explore a possible field of application such as digital instrumentation. We start this chapter with a plea for using fuzzy logic to process instrumentation data. Working side-by-side with a laboratory that builds innovative sensors brings a true playground for AI. We will present two achievements. The first one is the use of knowledge to create high-level features that improve the performance of various classifiers. The last one is the adaptation of fuzzy decision tree induction algorithms to consider imperfect data that are represented by probabilistic distributions.

All these contributions help ExpressIF[®] being an XAI based on fuzzy logic and symbolic AI. Indeed, it suits the definition of XAI as a system that can learn and explain its decisions.

Finally, the last part is conclusive and consists in a unique chapter that draws the conclusions and perspectives of this work.

The annexes of this manuscript give exhaustive information about my resume, my teaching activities, the collective responsibilities I undertook, the supervision of young engineers and researchers, the different projects I worked on, the complete list of my publications and a selection of few articles.

As I chose to gather the contributions regarding their topics, it does not reflect the temporality of the research activities. In particular, it does not show the bond between the doctorates of Olivier Hotel and Noëlie Cherrier. Indeed, in Olivier’s work we wanted to show that using knowledge to build features could increase the performances. At this stage, the features were built manually. In Noëlie’s work, we went a step further by automating the production of interpretable features. To remedy this problem, figure 1.7 presents the timeline of my research activities and the corresponding sections. It provides another way to read this manuscript.

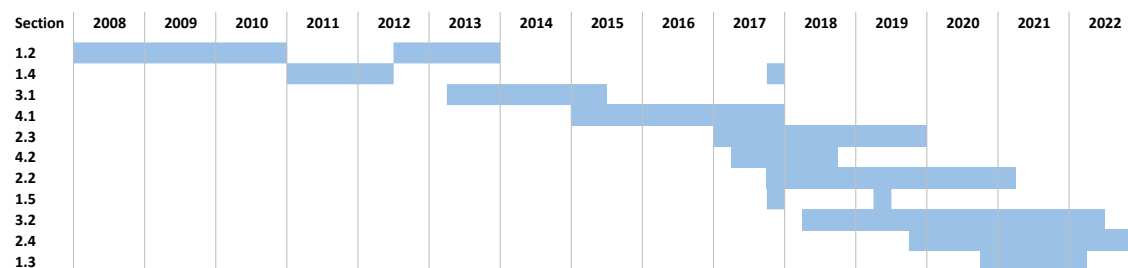


Figure 1.7: Research activities timeline.

Chapter 2

Background on Fuzzy Systems

2.1	Fuzzy sets	17
2.1.1	Operations on fuzzy sets	19
2.1.2	Cartesian product	23
2.2	Fuzzy relations	24
2.2.1	Operations on fuzzy relations	24
2.2.2	Specific properties	25
2.2.3	Similarity relations	25
2.2.4	Fuzzy spatial relations	25
2.2.5	Fuzzy mathematical morphology	26
2.3	Knowledge representation with fuzzy logic	27
2.3.1	Fuzzy granules	28
2.3.2	Linguistic variables	28
2.3.3	Fuzzy propositions, predicates and expressions	29
2.4	Reasoning in fuzzy logic	30
2.4.1	Fuzzy rules	30
2.4.2	Fuzzy implications	30
2.4.3	Generalized <i>Modus Ponens</i>	31
2.5	Fuzzy rule-based systems	33
2.5.1	Principles	33
2.5.2	Specific types of FRBS	34
2.6	Fuzzy Constraint Satisfaction Problems	36

In this chapter, I will define the key concepts to understand the contributions that are developed in the next part. It must be seen as a short tutorial that is not presenting the state-of-the-art in the discipline. The reader who is familiar with fuzzy logic may skip this chapter.

Fuzzy logic is a multi-valued logic that generalizes the classical logic to allow the representation of vague or imprecise knowledge and a soft reasoning on them. It is based on the concept of fuzzy sets that has been developed by L.A. Zadeh (Zadeh, 1965).

2.1 Fuzzy sets

Let \mathcal{U} be a reference set called *Universe of Discourse*. Let X be a subset of \mathcal{U} . In classical set theory, we can indicate which elements belong to X using a membership function:

$$\mu_X : \mathcal{U} \rightarrow \{0, 1\}$$

that maps each element $x \in \mathcal{U}$ with 1 if it belongs to X and 0 otherwise. Such a set will be called *crisp* in the remainder of the manuscript.

The introduction of fuzzy sets (Zadeh, 1965) brings a graduation of the membership of an item to a set. This allows defining concepts whose boundaries are not correctly defined (e.g. a hot temperature) or a gradual transition from one class to another (e.g. from a lukewarm to a hot temperature) (Bouchon-Meunier, 2007). It prevents from using rigid thresholds: for instance, if you are close to something within a distance of 2 meters, it is not natural not to be close at all by a distance of 2.01 meters.

Definition 2.1 (Fuzzy set)

A fuzzy set X of \mathcal{U} is defined by its membership function $\mu_X : \mathcal{U} \rightarrow [0, 1]$ that maps each element $x \in \mathcal{U}$ with the membership degree $\mu_X(x)$, i.e. the strength to which it belongs to X .

Thus, a crisp set is a particular fuzzy set whose membership function takes either 0 or 1 as values. It is common to represent a fuzzy set X as follows:

- if \mathcal{U} is finite, $X = \sum_{x \in \mathcal{U}} \frac{x}{\mu_X(x)}$,
- if \mathcal{U} is infinite, $X = \int_{x \in \mathcal{U}} \frac{x}{\mu_X(x)}$.

For instance, let A and B be two fuzzy sets, defined on $\mathcal{U} = \{V_1, V_2, V_3\}$ such as:

$$\begin{aligned} A &= \frac{V_1}{0.2} + \frac{V_2}{1} + \frac{V_3}{0} \\ B &= \frac{V_1}{0.4} + \frac{V_2}{0.6} + \frac{V_3}{0.8}. \end{aligned}$$

Different features are related to a fuzzy set X , as defined in the following.

Definition 2.2 (Height)

The height of a fuzzy set X is the maximum membership degree of its elements:

$$hgt(X) = \sup_{x \in \mathcal{U}} \mu_X(x).$$

Definition 2.3 (Support)

The support of a fuzzy set X is the crisp set containing all the elements in \mathcal{U} with a non-null membership degree:

$$supp(X) = \{x \in \mathcal{U} : \mu_X(x) \neq 0\}.$$

Definition 2.4 (Core)

The core of a fuzzy set X is the crisp set containing all the elements in \mathcal{U} with a membership degree equals to 1:

$$core(X) = \{x \in \mathcal{U} : \mu_X(x) = 1\}.$$

Definition 2.5 (Cardinality)

The cardinality of a fuzzy set X is the global membership degree of all its elements:

$$|X| = \sum_{x \in \mathcal{U}} \mu_X(x).$$

A *normalized fuzzy set* is a fuzzy set with at least one element whose membership degree is 1. Thus, for a normalized fuzzy set X , $hgt(X) = 1$, $core(X) \neq \emptyset$ and $supp(X) \neq \emptyset$.

Finally, there are two other ways to get a crisp set that matches approximately a fuzzy set, respectively the α -cut and strict α -cut.

Definition 2.6 (α -cut and strict α -cut)

The α -cut of a fuzzy set X (resp. strict α -cut), denoted $[X]_\alpha$ (resp. $[X]_{\underline{\alpha}}$) is the crisp set defined by:

$$[X]_\alpha = \{x \in \mathcal{U} : \mu_X(x) \geq \alpha\}$$

resp.

$$[X]_{\underline{\alpha}} = \{x \in \mathcal{U} : \mu_X(x) > \alpha\}.$$

Obviously, for a fuzzy set X , we have:

$$\begin{aligned} supp(X) &= [X]_{\underline{0}}(X) \\ core(X) &= [X]_1(X). \end{aligned}$$

For instance, let us consider the two fuzzy sets A and B defined above:

$hgt(A) = 1$	$hgt(B) = 0.8$
$supp(A) = \{V_1, V_2\}$	$supp(B) = \{V_1, V_2, V_3\}$
$core(A) = \{V_2\}$	$core(B) = \emptyset$
$ A = 1.2$	$ B = 1.8$
$[A]_{0.4} = \{V_2\}$	$[B]_{0.4} = \{V_1, V_2, V_3\}$
$[A]_{\underline{0.4}} = \{V_2\}$	$[B]_{\underline{0.4}} = \{V_2, V_3\}$
A is normalized.	B is not normalized.

For a fuzzy set X defined on a infinite universe of discourse \mathcal{U} , figure 2.1 represents its different characteristics.

2.1.1 Operations on fuzzy sets

The simplest operation on two fuzzy sets is the equality. Two fuzzy sets A and B of \mathcal{U} are equal if their membership functions are equal:

$$\forall x \in \mathcal{U}, \mu_A(x) = \mu_B(x).$$

We can also define the inclusion of A in B by expressing that all element in \mathcal{U} that belongs to A (to any extent), belongs to B with a degree that is greater or equal:

$$\forall x \in \mathcal{U}, \mu_A(x) \leq \mu_B(x).$$

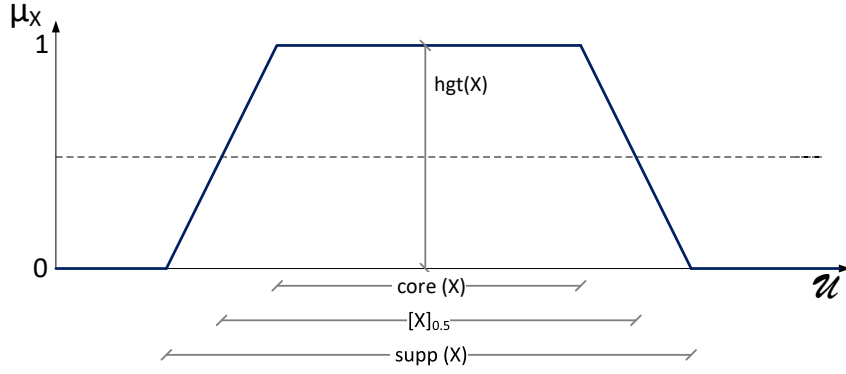


Figure 2.1: The different characteristics of a fuzzy set X .

The intersection and the union of two fuzzy sets have also been defined. However, their definitions are not unique and we consider instead families of operators.

Definition 2.7 (Triangular norm)

A triangular norm, or t -norm, is a function $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$ that satisfies the following properties for all $w, x, y, z \in [0, 1]$:

- *Commutativity*: $a \top b = b \top a$;
- *Associativity*: $\top(x, \top(y, z)) = \top(\top(x, y), z)$;
- *Monotonicity*: $\top(w, x) \leq \top(y, z)$ if $w \leq y$ and $x \leq z$;
- *Neutral element*: $\top(x, 1) = x$.

Table 2.1 presents the main norms. Their applications are illustrated in figure 2.2. They are used to defined the intersection between two fuzzy sets.

Definition 2.8 (Intersection)

Any t -norm is an intersection operator that builds a new fuzzy set $C = A \cap_{\top} B$ from two fuzzy sets A and B such that:

$$\forall x \in \mathcal{U}, \mu_C(x) = \top(\mu_A(x), \mu_B(x)).$$

Another family of operators are called triangular conorms.

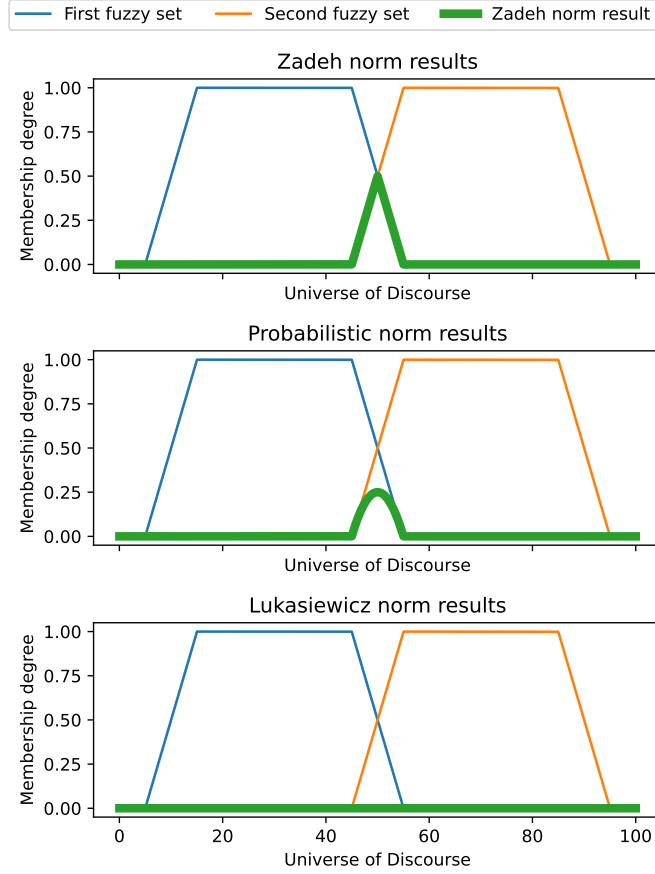


Figure 2.2: Examples of application of the main t-norms on two fuzzy sets.

Definition 2.9 (Triangular conorm)

A triangular conorm, or t-conorm, is a function $\perp : [0, 1] \times [0, 1] \rightarrow [0, 1]$ that satisfies the following properties for all $w, x, y, z \in [0, 1]$:

- Commutativity: $a \perp b = b \perp a$;
- Associativity: $\perp(x, \perp(y, z)) = \perp(\perp(x, y), z)$;
- Monotonicity: $\perp(w, x) \leq \perp(y, z)$ if $w \leq y$ and $x \leq z$;
- Neutral element: $\perp(x, 0) = x$.

Table 2.1 presents the main t-conorms. Their applications are illustrated in figure 2.3. This family of functions defines the union of two fuzzy sets.

Definition 2.10 (Union)

Any t-conorm is an union operator that builds a new fuzzy set $C = A \cup_{\perp} B$ from two fuzzy sets A and B such that:

$$\forall x \in \mathcal{U}, \mu_C(x) = \perp(\mu_A(x), \mu_B(x)).$$

Sometimes, t-norms (resp. t-conorms) are denoted \otimes (resp. \oplus). In this document,

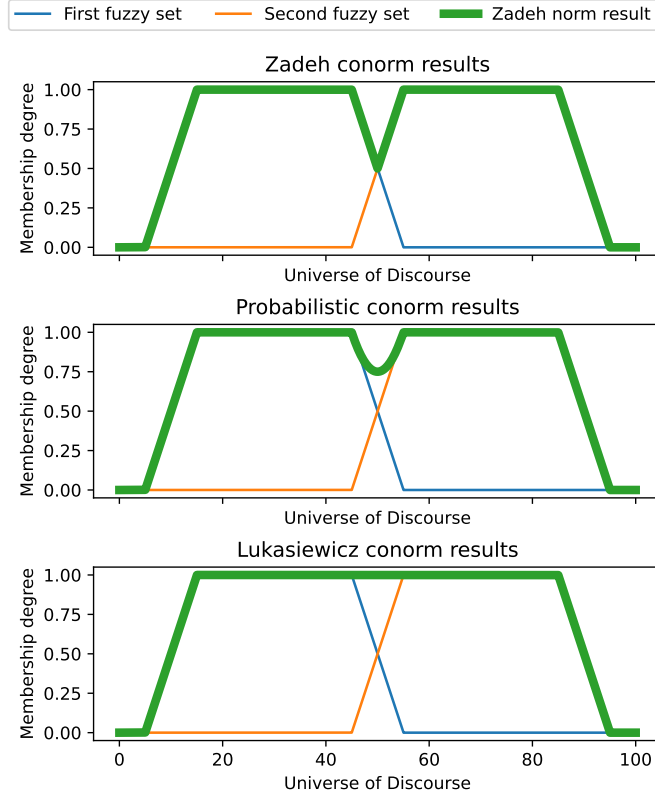


Figure 2.3: Examples of application of the main t-conorms on two fuzzy sets.

I will also borrow the symbols respectively \wedge and \vee from logics, when they are used as conjunctions and disjunctions respectively in logical formulas.

Different functions can define the complement of a fuzzy set and their family is called negations.

Definition 2.11 (Negation)

A negation is a function $n : [0, 1] \rightarrow [0, 1]$ that satisfies, for all $x, y \in [0, 1]$:

- $n(0) = 1$ and $n(1) = 0$;
- $n(x) \leq n(y)$ if $x \geq y$.

A negation is *strict* if it is continuous and if it satisfies $\forall x > y, n(x) < n(y)$. It is *involution* if and only if $\forall x, n(n(x)) = x$. As for t-norms and t-conorms, I will borrow the symbol \neg from logics to represent a strict negation when involved into an expression.

Definition 2.12 (Complement)

The complement X^{C_n} of a fuzzy set X of \mathcal{U} can be defined by an involutive negation such as:

$$\forall x \in \mathcal{U}, \mu_{X^{C_n}}(x) = n(\mu_X(x)).$$

Figure 2.4 illustrates the complement of a fuzzy set by the negation $n(x) = 1 - x$. A t-norm \top and a t-conorm \perp are said dual for a strict negation n if they satisfy, for

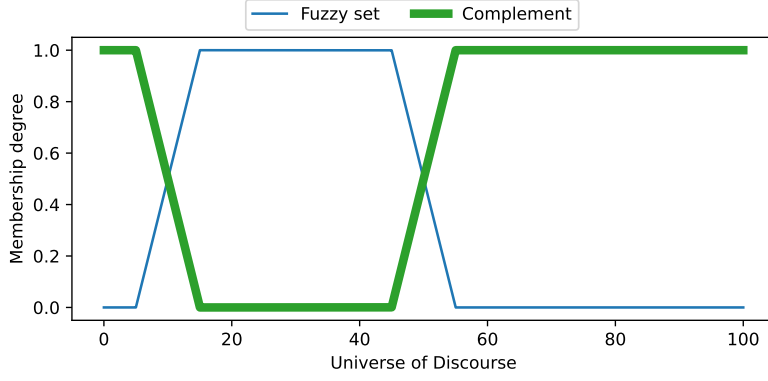


Figure 2.4: Example of the negation (complement) of a fuzzy set.

all $x \in [0, 1]$:

$$\begin{aligned} n(\top(x, y)) &= \perp(n(x), n(y)) \\ n(\perp(x, y)) &= \top(n(x), n(y)). \end{aligned}$$

The negation $n(x) = 1 - x$ allows the duality of the main t-norms and t-conorms (see table 2.1 for some examples).

Name	t-norm	t-conorm
Zadeh / Gödel	$\top_Z = \min(x, y)$	$\perp_Z = \max(x, y)$
Probabilistic	$\top_P = xy$	$\perp_P = x + y - xy$
Lukasiewicz	$\top_L = \max(x + y - 1, 0)$	$\perp_L = \min(x + y, 1)$

Table 2.1: Examples of dual t-norms and t-conorms.

In particular, a triplet $\langle \top, \perp, n \rangle$ is a De Morgan triplet if it satisfies $x \perp y = n(n(x) \top n(y))$.

2.1.2 Cartesian product

It is often useful to consider several reference universes at the same time, for instance for multi-criteria decision making. The global fuzzy set, whose components are the initial fuzzy sets, is their Cartesian product.

Definition 2.13 (*Cartesian product*)

Let $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ be universes of discourse and let \mathcal{U} be their Cartesian product $\mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_n$, whose elements are n -tuples (x_1, x_2, \dots, x_n) with $x_1 \in \mathcal{U}_1, x_2 \in \mathcal{U}_2, \dots, x_n \in \mathcal{U}_n$. From the fuzzy sets X_1, \dots, X_n , respectively defined on $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$, the Cartesian product $X = X_1 \times \dots \times X_n$ is a fuzzy set whose membership function is defined as:

$$\forall x = (x_1, \dots, x_n) \in \mathcal{U}, \mu_X(x) = \min(\mu_{X_1}(x_1), \dots, \mu_{X_n}(x_n)).$$

Let illustrate this definition by an example. Let A and B be two fuzzy sets defined as:

$$\begin{aligned} A &= \frac{V_1^A}{0.8} + \frac{V_2^A}{0.6} + \frac{V_3^A}{0.4} \\ B &= \frac{V_1^B}{0.3} + \frac{V_2^B}{0.7} \end{aligned}$$

then the Cartesian product $X = A \times B$ is the fuzzy set defined as:

$$X = \frac{(V_1^A, V_1^B)}{0.3} + \frac{(V_2^A, V_1^B)}{0.3} + \frac{(V_3^A, V_1^B)}{0.3} + \frac{(V_1^A, V_2^B)}{0.7} + \frac{(V_2^A, V_2^B)}{0.6} + \frac{(V_3^A, V_2^B)}{0.4}.$$

2.2 Fuzzy relations

Fuzzy relations are used to describe imprecise links (or gradual links) between elements.

Definition 2.14 (*Fuzzy relation*)

Let \mathcal{U}_X and \mathcal{U}_Y denote two universes of discourse. A fuzzy relation \mathcal{R} between \mathcal{U}_X and \mathcal{U}_Y is defined as a fuzzy set on $\mathcal{U}_X \times \mathcal{U}_Y$ whose membership function is

$$\mu_{\mathcal{R}} : \mathcal{U}_X \times \mathcal{U}_Y \rightarrow [0, 1].$$

If both \mathcal{U}_X and \mathcal{U}_Y are finite, a convenient way to define $\mu_{\mathcal{R}}$ is to use a matrix $M_{\mathcal{R}}$. For instance, let $\mathcal{U}_X = \{x_1, x_2, x_3\}$ and $\mathcal{U}_Y = \{y_1, y_2\}$, then a particular $\mu_{\mathcal{R}}$ can be defined as:

$\mathcal{U}_X \backslash \mathcal{U}_Y$	y_1	y_2
x_1	0.2	1
x_2	0	1
x_3	0.2	0.5

It is also possible to represent the relation as a heatmap, with colors ranging from white (representing 0) to black (representing 1). This is the convention that will be observed in this document, except in case of a specific indication.

2.2.1 Operations on fuzzy relations

There are operations that can be performed on fuzzy relations, like the inverse and the composition.

Definition 2.15 (*Inverse of a fuzzy relation*)

Let \mathcal{U}_X and \mathcal{U}_Y denote two universes of discourse. The inverse of the relation \mathcal{R} between \mathcal{U}_X and \mathcal{U}_Y is the fuzzy relation \mathcal{R}^{-1} between \mathcal{U}_Y and \mathcal{U}_X defined by:

$$\forall x \in \mathcal{U}_X, \forall y \in \mathcal{U}_Y, \mu_{\mathcal{R}^{-1}}(y, x) = \mu_{\mathcal{R}}(x, y).$$

Definition 2.16 (*Composition*)

Let $\mathcal{U}_X, \mathcal{U}_Y, \mathcal{U}_Z$ denote three universes of discourse. Let \mathcal{R}_1 be a fuzzy relation between \mathcal{U}_X and \mathcal{U}_Y , and \mathcal{R}_2 a fuzzy relation between \mathcal{U}_Y and \mathcal{U}_Z . The composition $\mathcal{R} = \mathcal{R}_1 \circ \mathcal{R}_2$ between \mathcal{U}_X and \mathcal{U}_Z is defined by:

$$\forall x \in \mathcal{U}_X, \forall z \in \mathcal{U}_Z, \mu_{\mathcal{R}}(x, z) = \sup_{y \in \mathcal{U}_Y} \min(\mu_{\mathcal{R}_1}(x, y), \mu_{\mathcal{R}_2}(y, z)).$$

This definition, called max-min composition, is the most used among many possibilities, in particular replacing the min by a t-norm.

2.2.2 Specific properties

As for classical relations, we can describe fuzzy relations \mathcal{R} defined on $\mathcal{U} \times \mathcal{U}$ with some properties:

- \mathcal{R} is *symmetrical* if $\forall (x, y) \in \mathcal{U} \times \mathcal{U}, \mu_{\mathcal{R}}(x, y) = \mu_{\mathcal{R}}(y, x)$.
- \mathcal{R} is *reflexive* if $\forall x \in \mathcal{U}, \mu_{\mathcal{R}}(x, x) = 1$.
- \mathcal{R} is *transitive* if $\mathcal{R} \supseteq \mathcal{R} \circ \mathcal{R}$. In particular, \mathcal{R} is *max-min transitive* if we use the max-min composition of fuzzy relations

$$\forall (x, z) \in \mathcal{U} \times \mathcal{U}, \mu_{\mathcal{R}}(x, z) \geq \sup_{y \in \mathcal{U}} \min(\mu_{\mathcal{R}}(x, y), \mu_{\mathcal{R}}(y, z)).$$

- \mathcal{R} is *anti-symmetric* if

$$\forall (x, y) \in \mathcal{U} \times \mathcal{U}, \mu_{\mathcal{R}}(x, y) > 0 \text{ and } \mu_{\mathcal{R}}(y, x) > 0 \Rightarrow x = y.$$

2.2.3 Similarity relations

A specific family of relations is used to characterize the similarity and is useful for classification problems.

Definition 2.17 (*Similarity relation*)

A similarity relation \mathcal{R} evokes the similarity of two elements of the universe of discourse \mathcal{U} and is symmetrical, reflexive and max-min transitive.

If \mathcal{R} is such that $\mu_{\mathcal{R}}(x, y) = 1$ iff $x = y$, a distance d can be paired with \mathcal{R} , such that:

$$\begin{aligned} d : \mathcal{U} \times \mathcal{U} &\rightarrow [0, 1] \\ (x, y) &\mapsto 1 - \mu_{\mathcal{R}}(x, y). \end{aligned}$$

2.2.4 Fuzzy spatial relations

Spatial relations are another specific family of relations. They play an important role in scene and image understanding (Bloch and Maitre, 1995; Bloch Isabelle, 2023). In this case, objects are represented as 2D fuzzy sets. For instance, in an image, the membership function that defines an object indicates which pixels belong to this object (eventually, partially).

Hence, it can be useful to represent the intersection of two objects. Since the classical intersection of fuzzy sets does not account for different overlapping situations, the following definition, which enables to get the notion of spatial overlapping, is preferred:

Definition 2.18 (*Fuzzy Spatial Degree of Intersection (Bloch, 2005)*)

For two fuzzy sets F and G defined on a universe \mathcal{U} with μ_F and μ_G as membership functions respectively, the fuzzy spatial degree of intersection is

$$\mathcal{R}_{int}(F, G) = \frac{V_H(\top(\mu_F, \mu_G))}{\min(V_H(\mu_F), V_H(\mu_G))}$$

with \top a t -norm and V_H the hypervolume of a fuzzy set defined as: $V_H(\mu) = \sum_{u \in \mathcal{U}} \mu(u)$.

A fuzzy degree of inclusion can also be derived from the crisp case.

Definition 2.19 (Fuzzy Degree of Inclusion (Bloch, 2005))

For two fuzzy sets F and G defined on a universe \mathcal{U} with respectively μ_F and μ_G as membership functions, the fuzzy degree of inclusion is

$$\mathcal{R}_{inc}(F, G) = \inf_{u \in \mathcal{U}} \perp(n(\mu_F(u)), \mu_G(u))$$

with \perp a t -conorm and n a fuzzy negation.

The literature defines many other fuzzy spatial relations, which can be split into three categories:

- *Topological relations* enable to express spatial configurations such as the adjacency between two regions or their overlapping;
- *Metric relations* rely on a measure. Directional relations characterize the orientation of an object relatively to a reference object. Metric relations are based on the many distances that have been proposed in the literature, even if most of them compare two membership functions and do not include any spatial information (Bloch, 2005);
- *Structural relations* describe a pattern between two or more objects such as *between* relation, *surrounds* relation. They can be seen as an extension of simpler directional relations between two objects.

There are different ways to define such relations. In particular, the fuzzy mathematical morphology (Bloch, 1999b) offers a theoretically powerful and well-defined framework.

2.2.5 Fuzzy mathematical morphology

Mathematical morphology relies on processing a set with another set called structuring element. The two main operators are the *erosion* and the *dilation*. Both of these operators have a fuzzy extension which can be obtained by a direct translation of crisp concepts into their fuzzy counterparts. Here, we focus on the fuzzy dilation.

Definition 2.20 (Fuzzy dilation (Bloch and Maitre, 1995))

Let \mathcal{U} be a universe of discourse, F and SE be fuzzy sets defined on \mathcal{U} , the fuzzy dilation of F by the structuring element SE is noted $D_{SE}(F)$ and defined as:

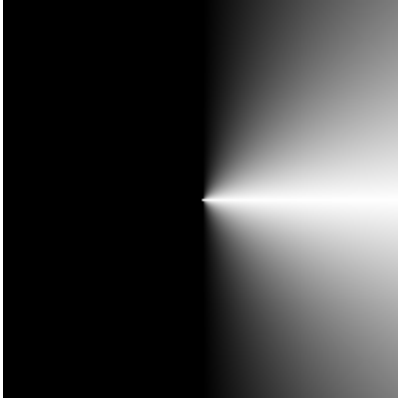
$$\forall u \in \mathcal{U}, D_{SE}(F)(u) = \sup_{v \in \mathcal{U}} \top(\mu_{SE}(u - v), \mu_F(v)).$$

In the context of directional relations, the structuring element is defined as:

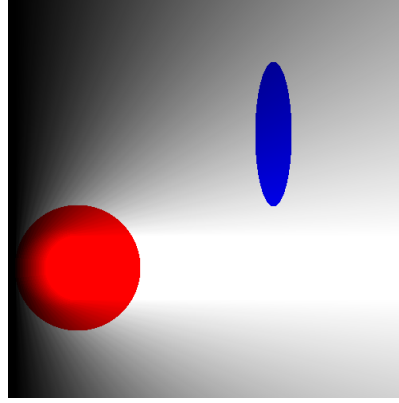
$$\forall u \in \mathcal{U}, \mu_{SE}^\theta(u) = \max \left[0, 1 - \frac{2}{\pi} \arccos \frac{\vec{u} \cdot \vec{v}_\theta}{\|\vec{u}\|} \right]$$

with \vec{u} the vector between the origin and u , and \vec{v}_θ the unit vector in the direction θ . Figure 2.5a shows an example of structuring element for $\theta = 0$, which represents the relation *to the right of*.

The result of the dilation can be visualized as a fuzzy landscape (Bloch, 2005). An example of fuzzy landscape is displayed on figure 2.5b. It is the result of the fuzzy dilation of the red disk by the structuring element that we can see on figure 2.5a. In this figure, a black point represent the value 0 and a white point a value of 1. This is not usual, but it allows to superimpose the objects and the fuzzy landscape.



(a) Example of structuring element for the relation *to the right of*.



(b) Example of fuzzy landscape.

Figure 2.5: Example of fuzzy dilation. Unusually, black stands for 0 and white for 1.

In figure 2.5b, in order to assess to which degree the blue ellipse is to the right of the red disk, we can use a fuzzy pattern matching approach and compute the necessity (the degree of inclusion) and the possibility (the degree of intersection) (Bloch, 1999b). A consequence of that process is that we do not need to compute another fuzzy landscape to evaluate the relation “to the right of the red disk” for any other entity in \mathcal{U} .

The fuzzy mathematical morphology (Bloch, 1999a) can also be used to model metric relations, such as “at a distance less than d ”, “at a distance greater than d ” and “at a distance between d_1 and d_2 ”. The degree to which a point $u \in \mathcal{U}$ is at a distance between d_1 and d_2 from the fuzzy object F is:

$$d(u, F) = \top(D_{SE_2}(F)(u), 1 - D_{SE_1}(F)(u))$$

D_{SE_1} the dilation by the structuring element SE_1 and D_{SE_2} the dilation by the structuring element SE_2 . SE_1 and SE_2 are defined as follows (Vanegas, 2011):

$$SE_1(u) = \begin{cases} 1 - \mu_n(d_E(u, O)) & \text{if } d_E(u, O) \leq d_1, \\ 0 & \text{otherwise.} \end{cases}$$

$$SE_2(u) = \begin{cases} 1 & \text{if } d_E(u, O) \leq d_2, \\ \mu_n(d_E(u, O)) & \text{otherwise.} \end{cases}$$

with μ_n the membership function of a fuzzy set whose core is $[d_1, d_2]$. O is the origin of the structuring element. Thus, d is actually the conjunction of a distance smaller than d_2 (D_{SE_2}) and a distance larger than d_1 ($1 - D_{SE_1}$).

(Cinbis and Aksoy, 2007) proposed another method, which is also based on mathematical morphology. For a directional relation, we saw that the directional information is brought by the term $\frac{2}{\pi} \arccos \frac{\vec{u} \cdot \vec{v}_\theta}{\|\vec{u}\|}$ of the structuring element defined in equation 2.1. For distances, instead of dealing with angular information, we can use the norm, such as:

$$\forall u \in \mathcal{U}, \mu_{SE}^\tau(u) = \max\left(0, 1 - \frac{\|\vec{u}\|}{\tau}\right)$$

with $\tau \in \mathbb{R}^{+*}$.

2.3 Knowledge representation with fuzzy logic

Fuzzy logic is a methodology for computing with words. It is a bridge between classical computing, i.e. with numbers and symbols, and the fact that humans employ words to

reason (Zadeh, 1996). In (Zadeh, 1997), Zadeh describes the human cognition as three features: granulation, organization and causation. Information granulation consists in decomposing a whole into significant parts, organization is about the arrangement of parts into a whole and finally, causation is searching for associations of causes with effects.

Hence, knowledge representation with fuzzy logic starts with the concept of *granule*.

2.3.1 Fuzzy granules

Definition 2.21 (*Information granule (Zadeh, 1996)*)

A granule is a fuzzy set of points that are arranged together due to their similarity, their indistinguishability or coherency. It is labeled by a linguistic term.

In general, perception is granular: time granules (for instance, the seconds, the years, etc.), image granules (e.g. regions, objects, etc.), etc. The strength of fuzzy logic is thus to represent granules with boundaries that are not precisely defined.

A granule can be seen as a constraint on a variable. For instance, in the proposition “Mary is young”, the granule *young* constrains the age of Mary (Zadeh, 1996).

2.3.2 Linguistic variables

Linguistic variables (Zadeh, 1975) are variables whose values are linguistic terms. They are a specific case of granulation of a universe of discourse \mathcal{U} .

Definition 2.22 (*Linguistic variable*)

A linguistic variable is a triplet (V, \mathcal{U}, T_V) where:

- V is a variable (e.g. age, temperature, etc.);
- \mathcal{U} is a universe of discourse on which V is defined;
- $T_V = \{A_1, A_2, \dots\}$ is a set (most of the time a finite set) of fuzzy sets of \mathcal{U} that are used to characterize V .

Each fuzzy set of T_V is a granule of \mathcal{U} : the granulation is more or less fine regarding the size of T_V . A linguistic variable offers a partition of \mathcal{U} . Usually, the A_i denote both the fuzzy set and the linguistic term.

There is a special case of partition, called strong partition, which meets the condition:

$$\forall x \in \mathcal{U}, \sum_{A \in T_V} \mu_A(x) = 1$$

Figure 2.6 shows an example of a linguistic variable “Temperature”, defined on $[-30^\circ C; 60^\circ C]$, with three terms “cold”, “average” and “hot”. It is a strong partition of its domain of definition.

The linguistic variables are generally used alongside with linguistic modifiers, for instance “very”, like in “very hot”.

Definition 2.23 (*Linguistic modifier*)

A linguistic modifier is an operator m that gives, from a fuzzy set A of a linguistic variable V , a new fuzzy set $m(A)$. Let t_m be the mathematical transformation attached to m , then $\mu_{m(A)} = t_m(\mu_A)$.

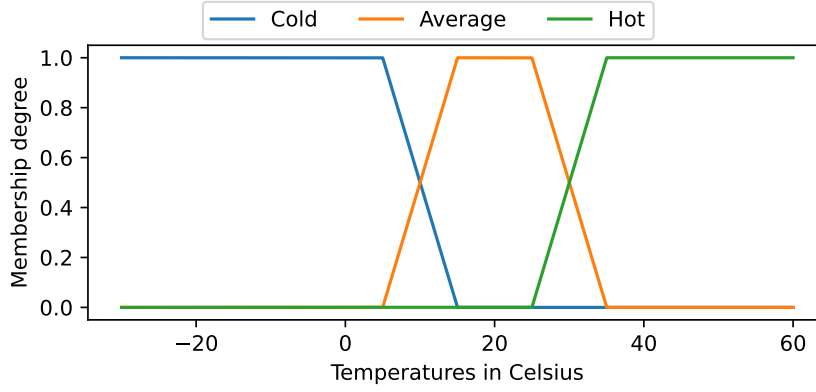


Figure 2.6: Example of a linguistic variable “Temperature”, defined on $[-30^{\circ}C; 60^{\circ}C]$, with three terms “cold”, “average” and “hot”.

Some modifiers strengthen the characterization, like “very” that Zadeh defined by $t_m(\mu_A(x)) = \mu_A(x)^2$, and others weaken it, like “more or less” that he defined by $t_m(\mu_A(x)) = \sqrt{\mu_A(x)}$. A negation “not” can also be considered as a modifier whose transformation is $t_m(\mu_A(x)) = 1 - \mu_A(x)$.

Let M be a set of modifiers and (V, \mathcal{U}, T_V) a linguistic variable, then $M(T_V)$ denotes the set of fuzzy characterizations obtained by the combinations of the modifiers of M on T_V . For instance, let $T_V = \{cold, average, hot\}$ and $M = \{very, not\}$, then $M(T_V)$ contains *very cold*, *very not hot*, *not average*, etc.

2.3.3 Fuzzy propositions, predicates and expressions

The representation of knowledge expressed symbolically (instead of numerically), uses either linguistic variables or fuzzy predicates.

Definition 2.24 (*Fuzzy proposition*)

Let L be a set of linguistic variables and M be a set of modifiers. A fuzzy proposition is defined from a linguistic variable (V, \mathcal{U}, T_V) of L and from the qualification “ V is A ” for a fuzzy set A belonging to T_V or $M(T_V)$. The truth value of a fuzzy proposition is given by the membership function μ_A of A .

Fuzzy propositions are sometimes referred to as fuzzy elementary propositions.

In this document, we also consider an elementary piece of knowledge called predicates.

Definition 2.25 (*Fuzzy predicate*)

A fuzzy predicate is a logical assertion that can contain variables and which takes values in $[0; 1]$ according to the values of these variables.

The fuzzy predicates generalize the concept of relations. They can also represent properties.

Finally, fuzzy expressions are obtained by the composition of other fuzzy expressions or fuzzy propositions. They are sometimes referred to as general fuzzy propositions. Let V and W be two linguistic variables.

A fuzzy expression can be:

- A negation, e.g. “ V is not A ”. In this case, the truth value is obtained with a negation n .

- A conjunction, e.g. “V is A and W is B”. In this case, the truth value is obtained by the application of a t-norm \top .
- A disjunction, e.g. “V is A or W is B”. In this case, the truth value is obtained by the application of a t-conorm \perp .

The combinations can be more complex. For instance, if U is a linguistic variable: “V is A and W is B or U is not C”. There is another way to combine fuzzy expressions: fuzzy rules, introduced in the next section.

2.4 Reasoning in fuzzy logic

The previous section showed how knowledge can be represented by fuzzy sets. Nevertheless, the classical logic cannot manipulate them to reason. Fuzzy logic is thus used to reason on imprecise, vague and eventually uncertain knowledge.

2.4.1 Fuzzy rules

In fuzzy reasoning, the fuzzy rule plays a preponderant part.

Definition 2.26 (Fuzzy rule)

A fuzzy rule represents a knowledge of the form “IF condition THEN conclusion” where condition is a fuzzy expression and conclusion may be a fuzzy proposition. They are linked together by a fuzzy implication.

The condition is also called the premise or the antecedent and the conclusion is also called the consequent.

For instance, a fuzzy rule can be “if V is A then W is B” or “if V is A and W is B then U is C”.

The value of the premise is sometimes referred as the rule activation.

2.4.2 Fuzzy implications

Any rule is paired with a fuzzy implication. It quantifies the strength of the link between the premise and the conclusion, to highlight the influence of the satisfaction of the premise on the satisfaction of the conclusion.

Definition 2.27 (Fuzzy implication)

A fuzzy implication \mathcal{I} associated with a fuzzy rule “if V is A then W is B” is defined from two universes of discourse \mathcal{U}_X and \mathcal{U}_Y .

The truth value of the implication is defined by the membership function $\mu_{\mathcal{I}}$ of a fuzzy relation between \mathcal{U}_X and \mathcal{U}_Y , such that:

$$\mu_{\mathcal{I}}(x, y) = \Phi(\mu_A(x), \mu_B(y))$$

with Φ a function such that, in the case of A and B are defined precisely and certainly, the fuzzy implication is identical to the implication of the classical logic.

To complete the definition, a fuzzy set A is defined precisely and certainly if its membership function μ_A takes the value 1 for a unique point, and 0 everywhere else.

There is not a unique way to generalize the classical implication. Table 2.2 shows the core fuzzy implications. In this table, there are two categories of implications: the first

ones respect the properties of the classical implication, whereas the second ones do not generalize the classical implication because they are conjunctions (Mamdani and Larsen).

Name	Notation	Truth value
Reichenbach	\mathcal{I}^R	$\mu_{\mathcal{I}^R}(x, y) = 1 - \mu_A(x) + \mu_A(x) \times \mu_B(y)$
Willmott	\mathcal{I}^W	$\mu_{\mathcal{I}^W}(x, y) = \max(1 - \mu_A(x), \min(\mu_A(x), \mu_B(y)))$
Rescher-Gaines	\mathcal{I}^{RG}	$\mu_{\mathcal{I}^{RG}}(x, y) = \begin{cases} 1 & \text{if } \mu_A(x) \leq \mu_B(y) \\ 0 & \text{otherwise} \end{cases}$
Kleene-Dienes	\mathcal{I}^{KD}	$\mu_{\mathcal{I}^{KD}}(x, y) = \max(1 - \mu_A(x), \mu_B(y))$
Brouwer-Gödel	\mathcal{I}^{BG}	$\mu_{\mathcal{I}^{BG}}(x, y) = \begin{cases} 1 & \text{if } \mu_A(x) \leq \mu_B(y) \\ \mu_B(y) & \text{otherwise} \end{cases}$
Goguen	\mathcal{I}^G	$\mu_{\mathcal{I}^G}(x, y) = \begin{cases} \min(\frac{\mu_B(y)}{\mu_A(x)}, 1) & \text{if } \mu_A(x) \neq 0 \\ 1 & \text{otherwise} \end{cases}$
Lukasiewicz	\mathcal{I}^L	$\mu_{\mathcal{I}^L}(x, y) = \min(1 - \mu_A(x) + \mu_B(y), 1)$
Mamdani	\mathcal{I}^M	$\mu_{\mathcal{I}^M}(x, y) = \min(\mu_A(x), \mu_B(y))$
Larsen	\mathcal{I}^P	$\mu_{\mathcal{I}^P}(x, y) = \mu_A(x) \times \mu_B(y)$

Table 2.2: Core fuzzy implications in a rule like “if V is A then W is B”. The two last implications do not generalize the classical implication.

Figure 2.7 shows the different values of the core implications, according to the values of $\mu_A(x)$ and $\mu_B(y)$. In particular, the plots highlight the differences between the core implications for the intermediary values of $\mu_A(x)$ and $\mu_B(y)$. They respect the conventions and map white with 0 and black with 1. As a quick reminder, the truth table of the implication in classical logic is:

p	q	$p \implies q$
0	0	1
0	1	1
1	0	0
1	1	1

Fuzzy implications that respect the truth values of the classical implication should thus have a value of 1 (black) on each corner of the plot, except on the bottom-right corner (i.e. for $\mu_A(x) = 1$ and $\mu_B(y) = 0$). The figure shows immediately that \mathcal{I}^M and \mathcal{I}^P do not respect that property and that they are conjunctions since the only value 1 is reached on the top-right corner, for $\mu_A(x) = \mu_B(y) = 1$.

2.4.3 Generalized *Modus Ponens*

Fuzzy logic offers equivalent reasoning capabilities as classical logic, in which the inference is performed by the *Modus Ponens*. In fuzzy logic, for a rule as “if V is A then W is B”, a conclusion must be obtained for W, in particular when the fact is “V is A’”, where A’ is more or less different from A. If A’ is close to A, i.e. their membership functions $\mu_{A'}$ and μ_A are little different, we expect B’ to be close to B. In particular, if A’ = A, we expect having B’ = B.

The concept of Generalized Modus Ponens (GMP) relies on Zadeh’s compositional rule of inference (Zadeh, 1973), which states, for a given relation \mathcal{R} defined on \mathcal{U}_V and \mathcal{U}_W , a fuzzy set A’ defined on \mathcal{U}_V , that the fuzzy set B’, induced by A’, is obtained from the composition of \mathcal{R} and A’: $B' = A' \circ \mathcal{R}$.

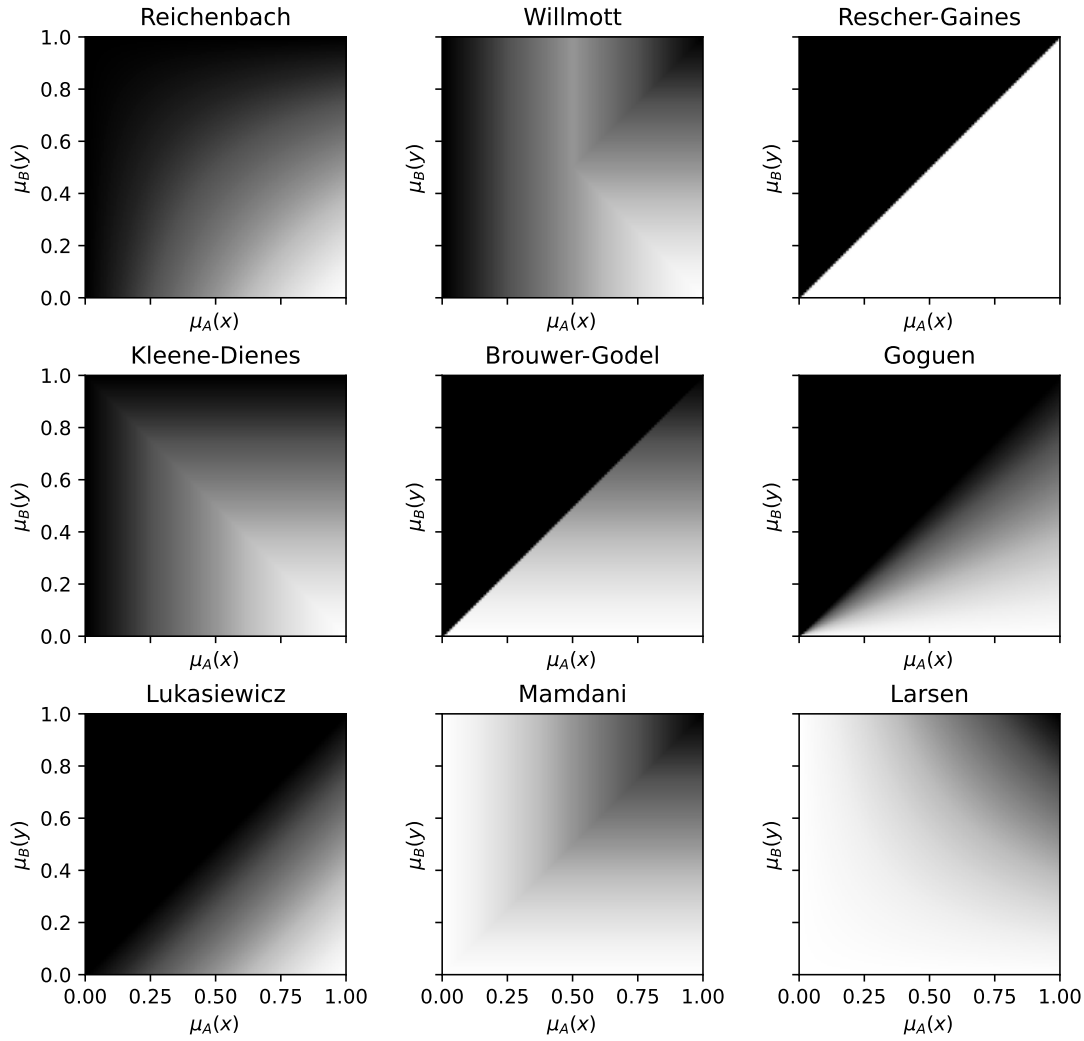


Figure 2.7: Plot of the truth values of the core implications according to $\mu_A(x)$ and $\mu_B(y)$.

Definition 2.28 (Generalized Modus Ponens)

The Generalized Modus Ponens is the inference operator for approximate reasoning. It processes in the same way symbolic and numeric data. Given two linguistic variables (V, \mathcal{U}_V, T_V) and (W, \mathcal{U}_W, T_W) , a rule in the form “if V is A then W is B ”, with fuzzy implication \mathcal{I} , and a fact “ V is A' ”, the GMP allows deriving the conclusion “ W is B' ”, whose membership function is:

$$\forall y \in \mathcal{U}_W, \mu_{B'}(y) = \sup_{x \in \mathcal{U}_V} \top(\mu_{A'}(x), \mu_{\mathcal{I}}(x, y))$$

where \top is a t -norm called GMP operator.

To ensure the compatibility with the classical *Modus Ponens*, some restrictions apply on the choice of the GMP operator regarding the choice of fuzzy implication. Table 2.3 shows the possible implications for the core t -norms.

Figure 2.8 shows different examples of GMP application, with the Zadeh GMP operator and the Brouwer-Gödel implication. The first row shows the rule “if V is A then W is B ”, with its premise A (left) and its conclusion B (right). The figure highlights some typical behaviors of the GMP:

GMP operator \top	Compatible implications \mathcal{I}
Lukasiewicz	$\mathcal{I}_R, \mathcal{I}_W, \mathcal{I}_{RG}, \mathcal{I}_{KD}, \mathcal{I}_{BG}, \mathcal{I}_G, \mathcal{I}_L, \mathcal{I}_M, \mathcal{I}_P$
Zadeh	$\mathcal{I}_{RG}, \mathcal{I}_{BG}, \mathcal{I}_M, \mathcal{I}_P$
Probabilistic	$\mathcal{I}_{RG}, \mathcal{I}_{BG}, \mathcal{I}_G, \mathcal{I}_M, \mathcal{I}_P$

Table 2.3: Compatibility between the core GMP operators and the fuzzy implications.

- The second row shows a fact A' that is included in A . In this case, the fact fulfill the premise, and we expect B' to be identical to B .
- The third row shows a fact A'' that is different from A . The GMP cannot infer the conclusion, so all the values of the universe of discourse are possible.
- The fourth row shows a fact A''' that is slightly different from A . The result of the GMP shows that more values are possible comparing to B .
- Finally, the last row shows a fact that includes A . The core of B is thus totally possible, but other values are uncertain.

2.5 Fuzzy rule-based systems

Fuzzy rule-based systems (FRBS) are one of the most important applications of fuzzy sets and fuzzy logic. They have been successfully applied to a large range of problems in different domains. FRBS are expert systems that use fuzzy sets and fuzzy logic as tools for representing knowledge and for inference. There exist different types of FRBS that share some common principles.

2.5.1 Principles

As any expert systems, FRBS have two main components: the knowledge base and the inference engine.

The knowledge base can be split into two parts: the vocabulary and the rule base. The vocabulary contains all the definition of granules and linguistic variables that partition the universes of discourse of the variables that are involved in the rule base. The rule base is simply a set of fuzzy rules as presented above. The rules represent the relationship between the inputs and the outputs of the FRBS.

The inference engine is in charge of the reasoning based on the application of the implications and GMP. The particularity of FRBS is that all the rules are evaluated at the same time, contrary to classical expert system where only one rule is fired at each step of the reasoning.

Let the rule base be composed of m rules

$$r_i: \text{if } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in} \text{ then } y \text{ is } B_i$$

with $i = 1, \dots, m$. Since several rules may contribute to the same output, it is necessary to aggregate the various rule results. We can distinguish two strategies:

- *First Aggregate, Then Infer* (FATI) consists in aggregating the different rules r_i as fuzzy relations into one overall fuzzy relation with the help of a fuzzy aggregation operator G , and then use the compositional rule of inference to get the final fuzzy set B' .

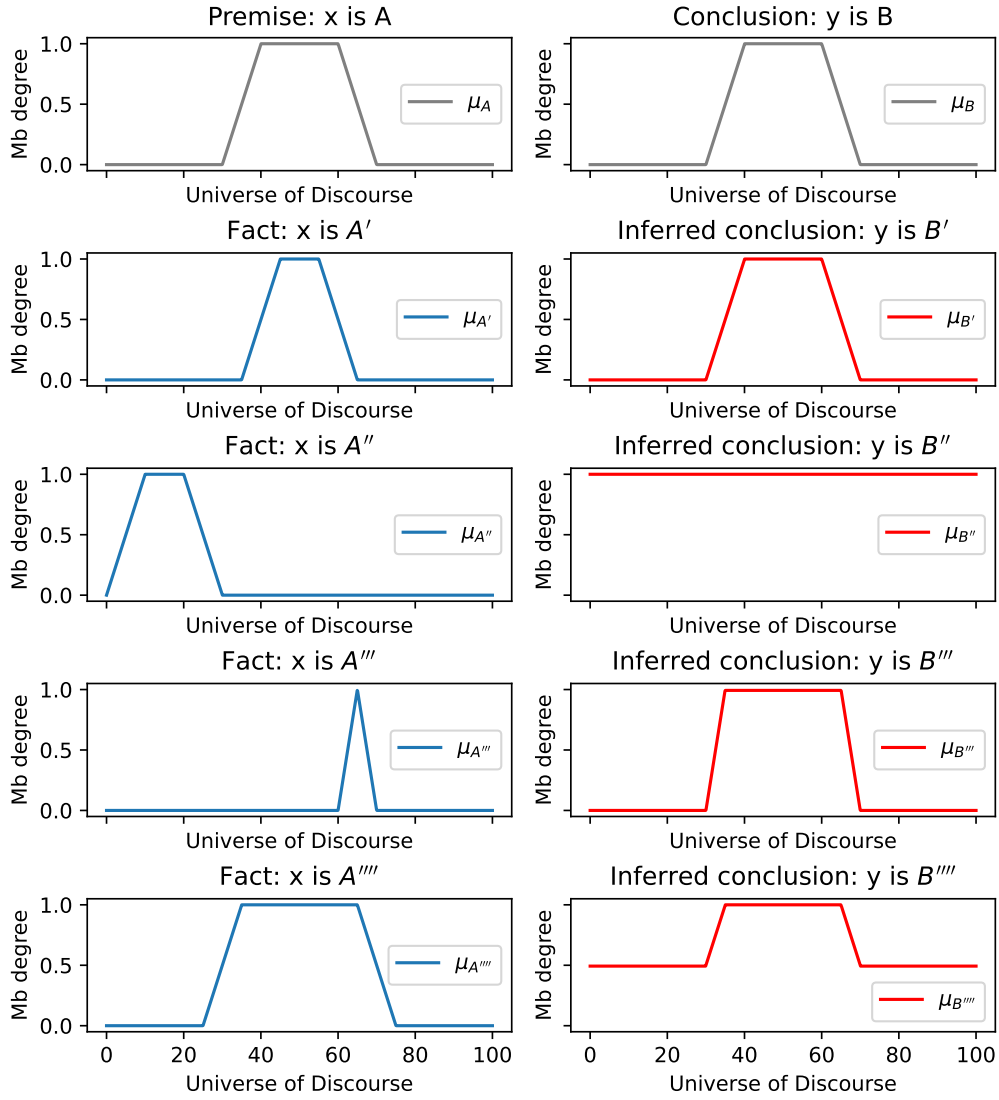


Figure 2.8: Different applications of GMP for $\top = \min$ and \mathcal{I}_{BG} .

- *First Infer, Then Aggregate* (FITA) consists in applying the GMP on each rule to generate m output fuzzy sets B'_i and then in aggregating them with a fuzzy aggregation operator G into one overall fuzzy set B' .

The FATI approach was the first proposed by Mamdani (Mamdani, 1974) but FITA is preferred, in particular for real-time systems, because it is simpler to implement and its computation is more efficient.

There exist different families of fuzzy aggregation operators, but the most used is the t-conorm, in particular max.

2.5.2 Specific types of FRBS

There exist different types of FRBS. In this section, we will just present three of them.

Mamdani fuzzy systems

Mamdani fuzzy systems have been introduced by Mamdani as fuzzy controllers. In a Mamdani fuzzy system, the inputs and the outputs are real numbers (and not fuzzy sets).

The first step aims at matching an input with a fuzzy set (i.e. a term of a linguistic variable): it is called *fuzzification*. Then the m rules

$$r_i: \text{if } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in} \text{ then } y \text{ is } B_i$$

are evaluated to get a fuzzy set B' . Since we need a real output, a last step is needed. It is called *defuzzification* and consists in obtaining one value from B' . Obviously, there are many defuzzification operators, and the most used is the center of gravity (CoG) and the mean of the maxima (MoM). Figure 2.9 shows the core defuzzification operators applied on the same fuzzy set, to highlight their differences.

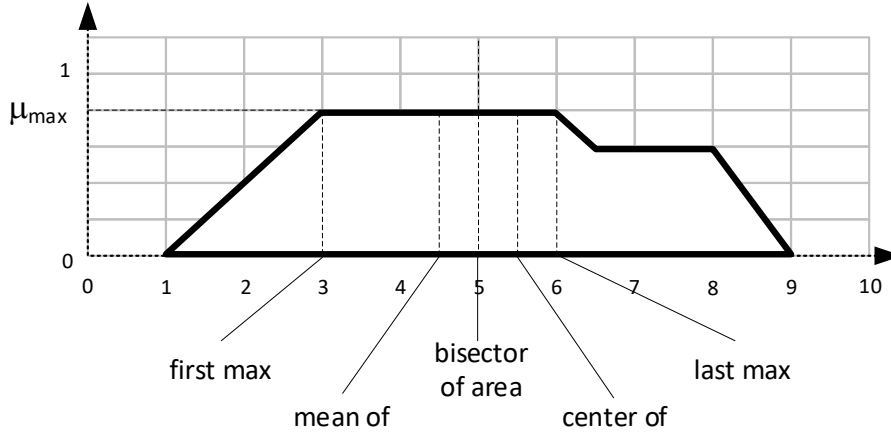


Figure 2.9: Illustration of the core defuzzification operators applied on a fuzzy set.

Takagi–Sugeno–Kang fuzzy systems

Takagi–Sugeno–Kang (TSK) fuzzy systems have also been introduced as fuzzy controllers. They differ from Mamdani systems by the form of the rule conclusions:

$$r_i: \text{if } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in} \text{ then } y = f_i(x_1, \dots, x_n)$$

where the $f_i(x_1, \dots, x_n) = \alpha_{i0} + \sum_{j=1}^n \alpha_{ij} \cdot x_j$ and the α_{ij} are real scalars, eventually null. In some extreme cases, a $f_i(x_1, \dots, x_n)$ can be a constant.

The real output can be obtained by the sum of the individual output values, weighted by the activation of their rule, and divided by the sum of the activations. As a consequence, the TSK fuzzy systems do not need a defuzzification step.

Fuzzy rule-based classifiers

A Fuzzy rule-based classifier (FRBC) is a particular type of fuzzy expert system that is intended to perform classification and in which rules have the form

$$r_i: \text{if } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in} \text{ then } y = c$$

with a crisp conclusion $c \in \mathcal{C}$ (the class) where \mathcal{C} is the set of all possible classes.

In its simplest form, the inference is performed with a *winner takes it all* approach: the class is given by the conclusion of the most activated rule. As a consequence, they do not need neither a fuzzy aggregation operator nor a defuzzification step.

2.6 Fuzzy Constraint Satisfaction Problems

A Constraint Satisfaction Problem (CSP) consists in assigning values to a set of variables that must respect a set of constraints. Dubois *et al.* (Dubois, Fargier, and Prade, 1996) present an extension of CSPs to the fuzzy logic framework to deal with imprecise parameters and flexible constraints: Fuzzy Constraint Satisfaction Problems (FCSP).

Definition 2.29 (*Fuzzy Constraint Satisfaction Problem*)

A Fuzzy Constraint Satisfaction Problem (FCSP) is defined by:

- A set of variables $X = \{x_1, \dots, x_m\}$.
- A set of domains $D = \{D_1, \dots, D_m\}$ such as D_i is the range of values that can be assigned to x_i .
- A set of flexible constraints $C = \{c_1, \dots, c_p\}$. A flexible constraint c_k is a pair (\mathcal{R}_k, S_k) where $S_k \subseteq X$ is the scope of the constraint, i.e. the variables that are involved in this constraint. \mathcal{R}_k is a fuzzy relation defined on $D_{k_1} \times \dots \times D_{k_m}$, the Cartesian product of domains of the variables in the scope of c_k . \mathcal{R}_k is defined by its membership function $\mu_{\mathcal{R}_k} : D_{k_1} \times \dots \times D_{k_m} \rightarrow [0, 1]$.

FCSPs allow to set preferences and priorities between the constraints, in particular to determine which constraints can be conflicted. As a consequence, it is necessary to be able to compare instantiations, notably to order them.

Definition 2.30 (*Consistency degree*)

Let γ be an instantiation $\{v_1, \dots, v_n\} \in D_1 \times \dots \times D_n$, the degree to which γ satisfies the FCSP is called the degree of consistency and is defined as the conjunction of the satisfaction of each of its constraints.

A FCSP is solved by algorithms that are adapted from classical CSP. The first one, backtracking algorithm, consists in few steps:

1. Select a variable that is not assigned yet;
2. Select a value for the chosen variable that satisfies the constraints, eventually partially;
3. Backtrack if no value can be assigned to the variable.

The backtracking algorithm is too naive to address large FCSPs. Dubois *et al.* (Dubois, Fargier, and Prade, 1996) also fuzzified the AC-3 algorithm into FAC-3. The algorithm acts like a preprocessing before the assignment of a value to a variable, deleting values from the domains of the variables that are not yet assigned. FAC-3 usually applies to problems with binary constraints (i.e. constraints with two variables in their scope). Vanegas *et al.* (Vanegas, Bloch, and Inglada, 2016) generalized FAC-3 to apply on any type of flexible constraint.

Now the main concepts of fuzzy systems have been defined, the next part describes my contributions.

Review of Research Activities

A person should set his goals as early as he can and devote all his energy and talent to getting there. With enough effort, he may achieve it. Or he may find something that is even more rewarding. But in the end, no matter what the outcome, he will know he has been alive.

Walter E. Disney

Chapter 3

Improving expressiveness of fuzzy systems

3.1	Expressiveness of AI systems	42
3.2	Online temporal predicates	43
3.2.1	Temporal scope	43
3.2.2	Examples	44
3.2.3	Base predicate and operator	45
3.2.4	Input variation predicates	48
3.2.5	Advanced predicates	49
3.2.6	Relation between <i>Ratio</i> and <i>Pers</i>	54
3.3	Spatial relations	54
3.3.1	<i>InTheDirection</i> relation	54
3.3.2	Definition of the ground level as a fuzzy landscape	58
3.4	Online spatio-temporal predicates	58
3.4.1	Base predicates	59
3.4.2	Trajectory predicates	60
3.4.3	Entrance and exit predicates	61
3.4.4	Compartmentalization predicates	62
3.4.5	Crossing predicate	63
3.5	Knowledge about the predicates	64
3.6	Applications	66
3.6.1	Crisis management	66
3.6.2	Water treatment plant management	69
3.7	Summary	69

Historically, expert systems were set up from expert knowledge. Cognitive engineers formalized this expertise into a formal logic language. This was a huge limitation of expert systems, especially when we aim at letting end-users handle them. To address this problem, it is useful to allow end-users expressing their knowledge with their own vocabulary.

With the introduction of linguistic variables, fuzzy logic brings a formalism that is closer to natural language and that allows modeling concepts that are too difficult to represent with numbers (Zadeh, 1996). The main principles and features of fuzzy logic are presented in chapter 2.

In this chapter, we tackle the problem of expressiveness that may prevent fuzzy systems from being used as XAI. Indeed, the earliest part of our research activities focused on

modeling new fuzzy predicates regarding different constraints:

- They must have a natural language representation: it will be helpful both to maintain the knowledge base interpretability and to generate a textual explanation. They represent properties of an object (e.g. *Round(.)*) or relations between objects (e.g. *ToTheLeftOf(.,.)*).
- They must be parsimonious and seamlessly configurable. This will ensure that they can be parameterized by end-users (directly or via an ergonomic graphical user interface) or by machine learning algorithms.
- It must be possible to evaluate their value from data. The fuzzy aspect allows considering the inaccuracy of the perception (e.g. sensors inaccuracy or image segmentation).

We consider fuzzy predicates as general as possible, i.e. higher-order predicates: the operands of the predicates may be functions, expressions, other predicates, etc.

First, we will define the expressiveness of such systems and then describe some of the predicates we developed (Iphar, Boudet, and Poli, 2021b; Le Yaouanc and Poli, 2012; Poli, Boudet, Espinosa, et al., 2017; Poli, Boudet, and Le Yaouanc, 2018; Poli, Boudet, and Mercier, 2016).

3.1 Expressiveness of AI systems

The term *expressiveness* (also referred to as *expressivity* or *expressive power*) has been first informally introduced in the 1980's in the field of knowledge engineering (Levesque and Brachman, 1987). Baader (Baader, 1996) proposes a formal – yet indirect – definition of the expressive power of knowledge representation languages: two languages have the same expressive power if and only if one language can be expressed by the other and reciprocally. Borgida (Borgida, 1996) extends this work to description logics and predicate calculus and bases the comparison on the meaning, i.e. all the possible interpretations of a given description or set of predicates.

More recently, the machine learning field has also used this terminology. Cohen *et al.* (Cohen, Sharir, and Shashua, 2016) defines the expressiveness of a neural network as the space of all possible configurations of parameters of the network. To Raghu *et al.* (Raghu et al., 2017), the expressive power of a neural network is linked to the influence of its architecture over the functions it computes.

As part of our work about XAI and fuzzy systems, we define expressiveness as follows.

Definition 3.1 (*Expressiveness*)

The expressiveness of an XAI is related to the vocabulary that can be used both in the knowledge base (e.g. rules, constraints) and in the explanations.

The expressiveness reflects the variety of situations that can be described by the XAI regarding the problem that it must solve.

The latter definition leads to the definition of *vocabulary*. In fuzzy systems, the vocabulary often designates the linguistic variables and their terms. In this work, the definition differs (see chapter 2, section 2.3 for the basics of knowledge representation in fuzzy logic).

Definition 3.2 (*Vocabulary*)

The vocabulary \mathcal{V} is a set of fuzzy propositions, relations and predicates that can be used by the XAI.

The motivation for this work comes from two postulates:

1. A richer vocabulary leads to a more expressive system, which should help to produce better decisions and explanations.
2. An XAI should rely on the same vocabulary a human would use for performing and describing the same task.

To enrich the vocabulary that can be used by our system, ExpressIF[®], we defined several predicates in different domains.

3.2 Online temporal predicates

The temporal predicates were introduced during the MobiSIC project (2007-2010), which led to the development of ExpressIF[®], during my postdoctoral fellowship.

This work is related to information streams produced by a set of sensors and how an expert can automate the decision process regarding the values and their changes over time. Formally, we consider ordered pairs (t, d) where t is a timestamp and d is data. Sensors may produce either measurements at a constant or a dynamic sampling rate (data streams, e.g. thermometers), or events when they occur (event streams, e.g. presence sensors).

Whatever the type of items in the streams, i.e. either data or events, the information may be incomplete and imprecise by nature (Artikis et al., 2014). A typology of temporal data imperfections has been proposed by Achich *et al.* (Achich et al., 2019). For instance, sensors may be out of order or inaccurate, or data may be noisy. For the last decade, an effort has been made to handle uncertainty in information flows in several ways (Alevizos et al., 2015). We can distinguish the contributions according to the ways authors consider the uncertainty: at the data source level, at the complex events description level or at the output level. Fuzzy logic is a good way to handle the uncertainty in all those levels, in particular for the vagueness of time, because of either the imprecision of time measurements or the subjective perception of time (Bouchon-Meunier, 2021).

Let us assume that the sensors give correct timestamps: the fuzziness is thus not necessary on timestamps but vagueness is useful to describe the possible relations between them. In addition, the different values from the sensors are fuzzified to both manipulate linguistic terms and handle their inaccuracy.

In the next section, I will define the concept of scope, which is necessary to compute the temporal predicates. I will introduce the example that will be used in this section. I will then present the base relations, and then the derivative ones.

3.2.1 Temporal scope

To ensure an effective implementation of the fuzzy temporal predicates, they are applied on operands that are fuzzy expressions, regarding a certain temporal scope, i.e. a limit in the historical data that are considered.

Definition 3.3 (Scope)

A scope S is a range over the past times that ends at the present moment. More formally, a scope is a fuzzy set defined on a temporal domain, whose membership function is used to weight the different past values and to define a vague notion of past, for example "recently" or the "last 5 seconds".

Figure 3.1 shows two examples of scopes, respectively representing "the last 5 seconds" and "the last 20 seconds".

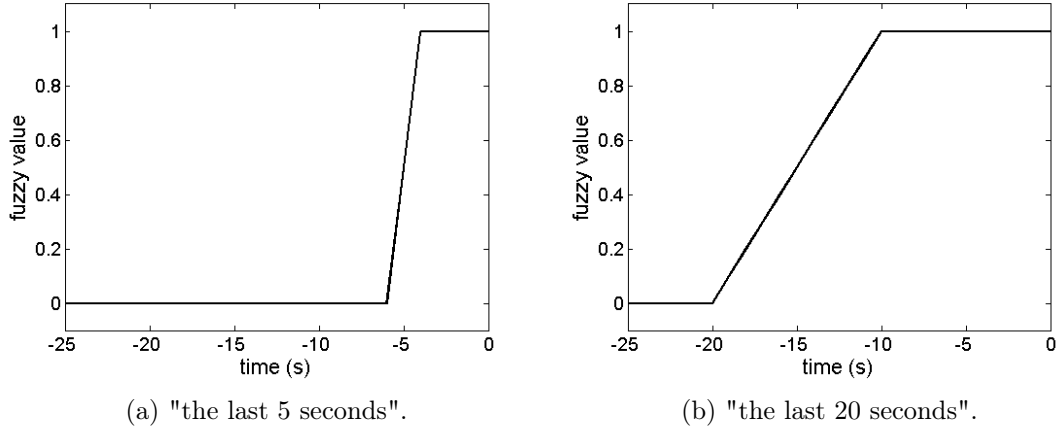


Figure 3.1: Membership functions representing two different temporal scopes.

The scope "the last 5 seconds" considers completely the 5 last seconds, and the events between 6 seconds and 5 seconds are considered with more and more importance. In other words, an event that occurred 5.5 seconds ago will influence the degree of fulfillment of the relation less than an event that occurred 3 seconds ago. It is used to add a security margin on the regular scope and avoids the threshold effect.

The scope "the last 20 seconds" has a different meaning. In the first 10 seconds of the scope, the event are totally considered, while in the last 10 seconds, they are gradually considered. This kind of scopes are intended to give different importance to the past events.

Theoretically, the scope may be infinite. However, for the sake of performance, it has to be finite. In practice, the membership function has a multiline shape, often semi-trapezoidal as shown in figure 3.1.

In terms of implementation, it is possible to determine a piecewise sampling rate according to the shape of both signal and membership function to guarantee a good precision of the predicates value. On a portion of the signal, if both the signal and the membership function are constant, then a low sampling rate is sufficient to guarantee a good accuracy in the computation of the operators.

Let S denote a temporal scope. As S is always relative to the present t_{now} , let $\mu_S(t - t_{now})$ be the value of the membership function of S for time t : to facilitate the reading, we will use the notation $\mu_S(t)$. S always contains the value at time t_{now} , and we denote S^* the scope $S \setminus \{t_{now}\}$.

3.2.2 Examples

Figure 3.2a shows an example of a signal for a temperature measurement. Regarding the fuzzy set "high" (figure 3.2b), figure 3.2c shows the signal for the proposition "the temperature is high". The temperature signal has been chosen to illustrate the usefulness and the behavior of the proposed temporal relations. This signal can be split into two parts:

- from the 5th to the 30th second, it describes cycles of high temperatures;
- from the 40th to the 55th second, it describes a steady high temperature state.

Consequently, the fuzzy expression "the temperature is high" catches these particular patterns with very high fuzzy values, and decreases to 0 between these two parts during a little bit less than 10 seconds.

Figure 3.3 shows another fuzzy proposition "the rainfall is high" that will be used in the remainder of this article. Note that both signals have a constant sampling rate of 1

second over 1 minute. After one minute, no more inputs are received from the sensors since the values are constant: this behavior is typical of event streams.

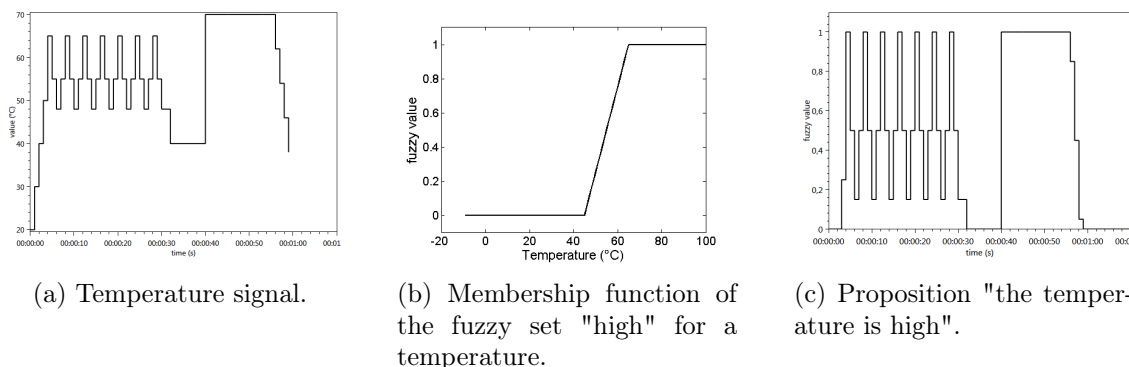


Figure 3.2: Input signal of a temperature sensor, a membership function associated to its domain, and evaluation of the resulting fuzzy proposition.

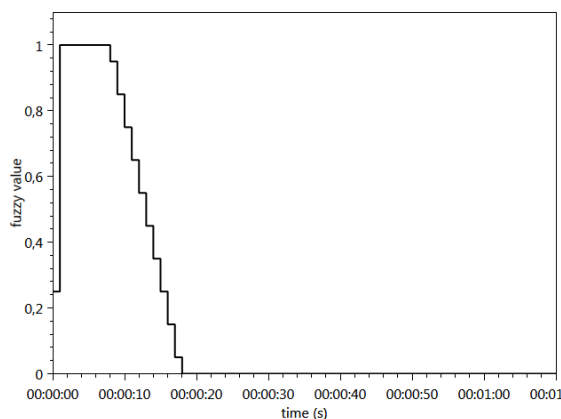


Figure 3.3: The fuzzy proposition "the rainfall is high".

Without loss of generality, the examples are given with Zadeh's t-norm and t-conorm.

3.2.3 Base predicate and operator

In this section, we describe the base predicate and operator that will be combined later into new ones.

Occurrence

The first fuzzy temporal predicate is unary and indicates whether a certain phenomenon has occurred regarding a given scope. Here a phenomenon is characterized by a fuzzy expression operand, like the fuzzy proposition "temperature is high", a conjunction "temperature is high and luminosity is high", or any fuzzy-valued expression. This predicate can be reformulated as: at least for one moment of the scope, the operand has a non-zero fuzzy value. It matches the specifications of the existential quantifier in (Barro et al., 2008) or the non-persistence relation in (Cariñena et al., 2000): it consists in the supremum, over all the moments t_i of the scope, of the conjunction between the value of the operand at time t_i and the membership function.

Definition 3.4 (Occurrence predicate)

Let E be a fuzzy expression, S a scope. The Occurrence predicate, for instance " E has occurred during S ", can be written:

$$Occ(E, S, t_{now}) = \sup_{t \in \text{supp}(S)} (E(t) \wedge \mu_S(t)). \quad (3.1)$$

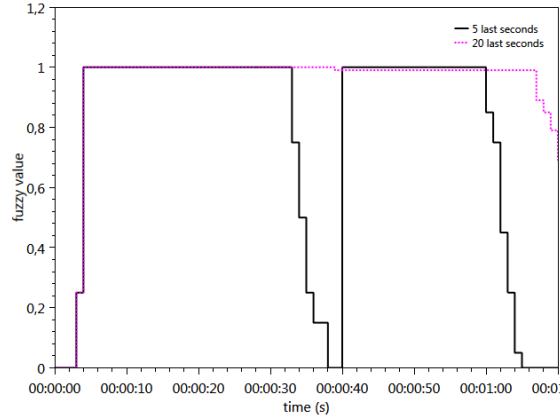


Figure 3.4: Application of the Occ operator to the expression "the temperature is high" on two different scopes.

Figure 3.4 shows the application of the Occ operator to the expression "the temperature is high" regarding both temporal scopes shown in figure 3.1. Whereas both curves raise identically at the beginning, the expression evaluated with the shortest scope decreases gradually to zero from the 34th to the 39th second: indeed, fuzzy values become lower and lower while older and older. On the contrary, with the largest scope, the output signal of the occurrence expression is smoother, hiding the gap obtained before.

Ratio operator

In the terminology of Cariñena *et al.* (Cariñena et al., 2000), we introduced a reduction operator called "ratio" that aggregates the different degrees of fulfillment of the operand expression E over a scope S . The result is then divided by a value corresponding to the sum of 1 over the same scope.

Definition 3.5 (Ratio operator)

Let E be a fuzzy expression, S a scope. The Ratio operator indicates how much an expression has been true over a scope. It can be formulated by:

$$Ratio(E, S, t_{now}) = \frac{\int_{t \in \text{supp}(S)} E(t) \wedge \mu_S(t)}{\int_{t \in \text{supp}(S)} \mu_S(t)} \quad (3.2)$$

where $\int_{t \in \text{supp}(S)} E(t) \wedge \mu_S(t)$ is the area under the signal composed of the past values of E regarding S and $\int_{t \in \text{supp}(S)} \mu_S(t)$ is the area under the membership function of S .

This operator is simply Zadeh's relative count applied to a temporal domain (Zadeh, 1983).

The figure 3.5 shows the application of the $Ratio$ operator to the proposition "the temperature is high" with both temporal scopes. As this operator is built on an integral,

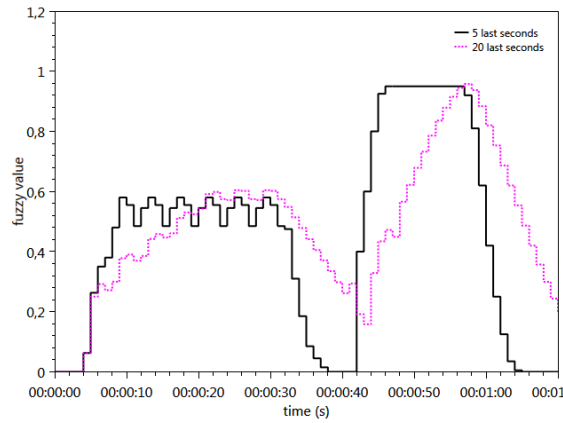


Figure 3.5: Application of *Ratio* operator to the expression "the temperature is high" on two different scopes.

it shows how it smooths the input signal. The larger the scope, the smoother the curve: larger scopes weaken the resulting signal slopes.

Expiration

The occurrence predicate and the ratio operator have a particularity when applied on a finite scope. Since our system has to deal with both event streams and data streams with dynamic sampling, inputs may not change for a while. Meanwhile, time goes by and the value of expressions containing *Occ* or *Ratio* may change. Their values can thus expire, that is to say that they are valid only for a while. As the scope has a finite duration and is anchored on t_{now} , they must be re-assessed.

Definition 3.6 (*Expiration*)

Expiration is a mechanism that allows pointing expressions in the rule base, whose values may not be valid anymore and must be reassessed in order to keep the coherence of the outputs. Expiration allows handling event streams or data streams with dynamic sampling rate.

The delay of expiration is customizable to fit the application of these operators. For instance, if events arrive every 15 minutes, it is not necessary to make them expire every second. This is a particular feature of ExpressIF[®]: the value of the outputs can change even if inputs have not changed. Expiration can also concern inputs: for instance, we can force an input whose value is given by an end-user to expire, forcing the end-user to enter a new value for it.

The expiration of *Occ* and *Ratio* is triggered regarding different criteria:

- The occurrence predicate keeps expiring as long as its value is strictly greater than 0. Indeed, if it reaches 0 and if operands do not change anymore, the occurrence will keep a null value, so it can stop expiring.
- The ratio operator $Ratio(E, t_{now}, S)$ keeps expiring if, at time t_{now} :
 - the signal associated with the scope S contains only one value;
 - $Ratio(E, S, t_{now})$ is null but $E(t_{now}) > 0$.

The effect of the expiration can be seen on figure 3.4 or on figure 3.5: the new expressions continue to be evaluated after one minute whereas input data did not change.

This next subsections aim at showing how the previous predicates can be combined into more complex temporal predicates that can be used conveniently by end-users.

3.2.4 Input variation predicates

In the case of the monitoring of complex systems, it is useful to make decisions regarding the variations of one input over a scope. We focus here on three types of variations, whether the values of an input is increasing, decreasing or just varying.

Growth and decline predicates

We defined the growth and the decline predicates to express this form of knowledge:

"input I increases/decreases <modifier> throughout S".

They are based on the gradient of the values of the input: $\nabla I(t)$. More precisely, the angle α between the gradient and the horizontal axis is used to characterize the growth or the decline of input value, as shown in Figure 3.6.

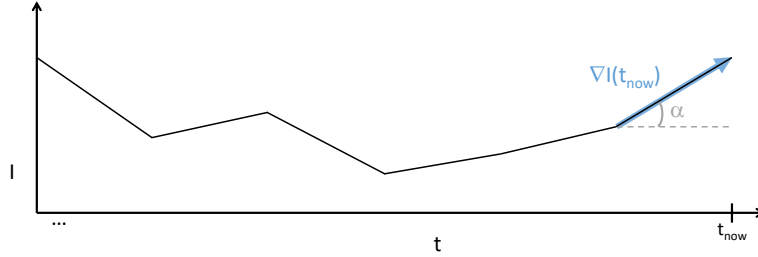


Figure 3.6: Principle of using the gradient to characterize growth or decline.

They also both use a modifier regarding α that helps characterizing the speed of the growth or decline. This modifier is a fuzzyset that is defined on $]0; \frac{\pi}{2}[$ for growth (resp. $] - \frac{\pi}{2}; 0[$ for decline). In other words, α is defined on the upper left quadrant of the trigonometric circle for growth, and on the bottom right quadrant for decline. To aggregate the values throughout the fuzzy scope, the *Ratio* predicate is used.

Definition 3.7 (*Growth/decline predicates*)

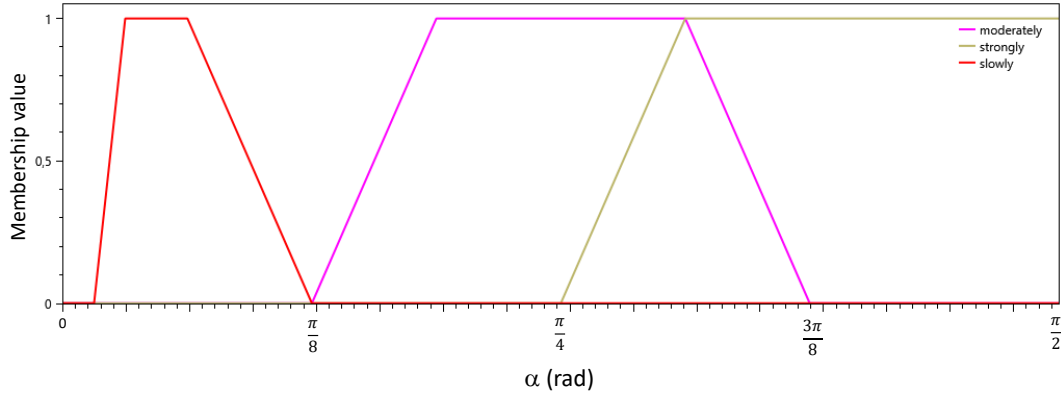
Let I be a crisp input with numerical values, $I(t)$ a signal of its past values, S a scope, M a valid modifier, i.e. a fuzzy set defined on the values of the angle α . The *Increases* and *Decreases* predicates can be written:

$$\text{Increases/Decreases}(I, S, M, t_{now}) = \text{Ratio}(\mu_m(\nabla I(t_{now})), S, t_{now}) \quad (3.3)$$

where μ_m is the membership function of the modifier fuzzy set M .

Figures 3.7 and 3.8 show respectively examples of modifiers for growth and decline predicates. Figure 3.7a (resp. Figure 3.8a) shows a fuzzy set defined on $]0; \frac{\pi}{2}[$ (resp. $]0; -\frac{\pi}{2}[$) and 3 different membership functions. Figures 3.7b, 3.7c and 3.7d (resp. 3.8b, 3.8c, 3.8d) show their respective geometric interpretations. A white pixel stands for the value 0 and a black pixel stands for the value 1.

It is important to note that these predicates only indicate if step by step, the input increases or decreases: they do not rely on the trend of the variations. Hence, if the



(a) Examples of modifiers for growth.

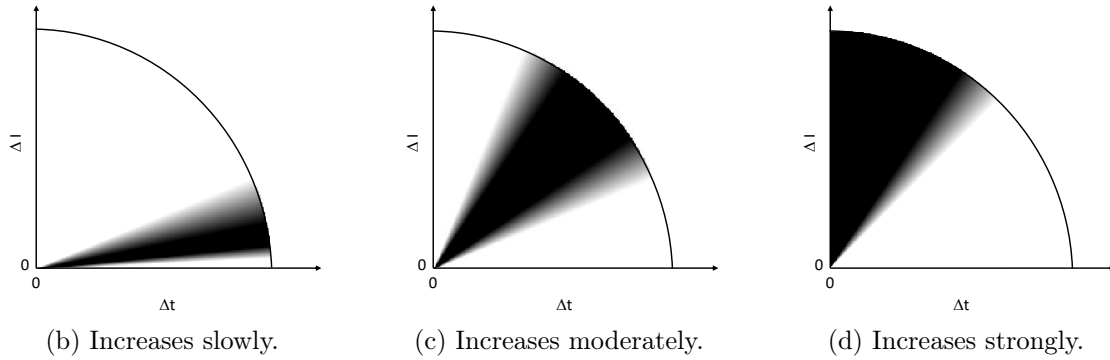


Figure 3.7: Examples of growth modifiers and their interpretations.

input signal $I(t)$ decreases (resp. increases) for a while, the *Increases* predicate (resp. *Decreases* predicate) will not fall down to 0. It is quite straightforward to change this default behavior, and we will discuss about that later in section 3.2.6.

Fluctuation predicate

This predicate measures how much an input varies throughout a fuzzy scope: "input I varies <modifier> throughout S ".

As seen previously, the modifier allows customizing the predicate and applies on the variance of I . The ratio operator is used to aggregate the values throughout the fuzzy scope.

Definition 3.8 (*Varies predicate*)

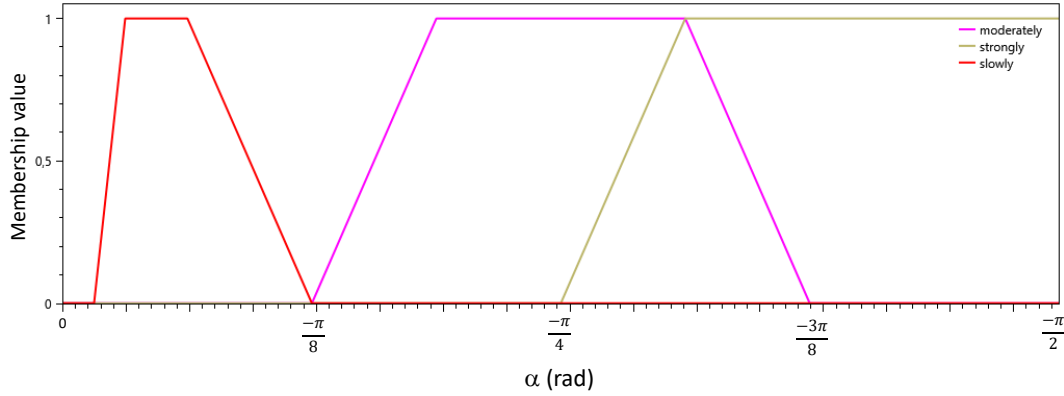
The *Varies predicate* is defined as:

$$\text{Varies}(I, S, M, t_{\text{now}}) = \text{Ratio}(\mu_m(\text{Var}(I, \text{supp}(S))), S, t_{\text{now}}) \quad (3.4)$$

where M is the modifier and μ_m its membership function, and $\text{Var}(I, \text{supp}(S))$ is the variance of I over the support of S .

3.2.5 Advanced predicates

In the previous section, we claimed the operand expression E can be any fuzzy-valued expression, including temporal expressions. The advantage is that the base predicate and



(a) Examples of modifiers for decline.

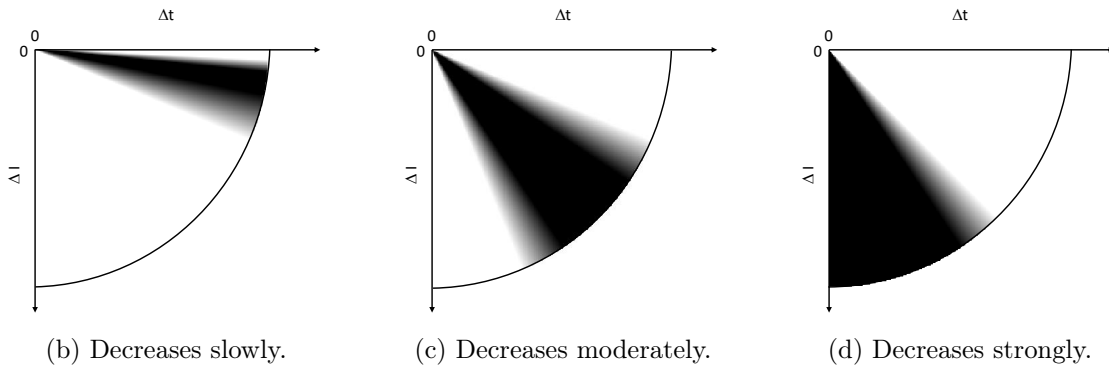


Figure 3.8: Examples of decline modifiers and their interpretations.

operator handle both the computation and the implementation of the expiration, so that the implementation of new predicates is straightforward.

We present further a few examples of advanced fuzzy temporal predicates, which are convenient to characterize the persistence or the precedence of the values of fuzzy expressions.

Persistence

We proposed a formalization of this concept based on the "*Occ*" operator. Indeed, the persistence of a phenomenon throughout a given scope is the fact that at each moment of the scope, the phenomenon is observed. In other words, for a given operand expression E , its negation must not have occurred.

Definition 3.9 (*Persistence predicate*)

Let E be a fuzzy expression and S a scope. The persistence of an expression throughout a fuzzy scope is defined by:

$$Pers(E, S, t_{now}) = \neg Occ(\neg E, S, t_{now}). \quad (3.5)$$

Figure 3.9 shows the application of the *Pers* operator to the fuzzy expression "the temperature is high" regarding both temporal scopes shown in figure 3.1. We can see that a persistence expression reaches its maximum value only when the steady state of high temperatures is itself reached, within a small delay.

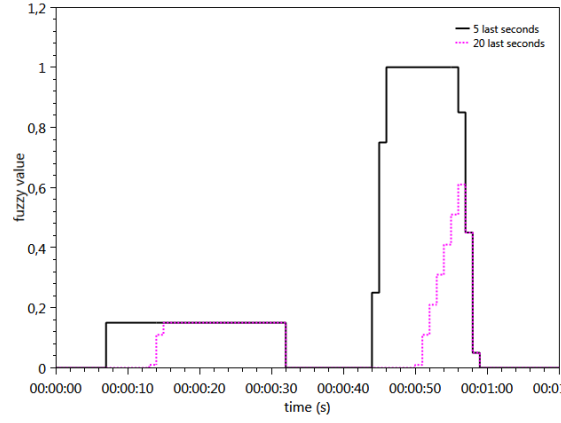


Figure 3.9: Application of the *Pers* operator to the expression "the temperature is high" on two different scopes.

This operator can perform only on signals with at least two samples to ensure whether the expression persists or not. Thus, it is possible to observe a small delay before the first evaluation of this operator.

We can observe the following relations between the predicates *Occ*, *Pers* and the *Ratio* operator:

$$\begin{aligned} Pers(E, S, t_{now}) &\leq E(t_{now}) \leq Occ(E, S, t_{now}) \\ Pers(E, S, t_{now}) &\leq Ratio(E, S, t_{now}) \leq Occ(E, S, t_{now}). \end{aligned} \quad (3.6)$$

With the *Occ* and *Pers* operators, we can investigate if a phenomenon has persisted for a while during the scope: for instance, "*E* has persisted at least 5 seconds during the last hour". Two scopes are involved in this relation: the first one characterizes the duration of the persistence, the second one the duration of the global relation. This kind of temporal relation can be very useful for system diagnosis and maintenance for instance.

The idea is to compute the persistence of a phenomenon on a scope $S_{duration}$, the last "5 seconds" in the example above. We then just have to look for the occurrences of such a persistence.

Definition 3.10 (*Persistence for a while*)

Let *E* be a fuzzy expression, *S* and $S_{duration}$ scopes. The persistence of a phenomenon for a while is given by:

$$PersWhile(E, S_{duration}, S, t_{now}) = Occ(Pers(E, S_{duration}, t_{now}), S, t_{now}). \quad (3.7)$$

By definition of the persistence operator, if the phenomenon disappears during $S_{duration}$, the *PersWhile* predicate is null. It is possible to be more tolerant by replacing the persistence by the *Ratio* operator.

Definition 3.11 (*Tolerant persistence for a while*)

Let *E* be a fuzzy expression, *S* and $S_{duration}$ scopes. The tolerant persistence of a phenomenon for a while is given by:

$$PersWhileTol(E, S_{duration}, S, t_{now}) = Occ(Ratio(E, S_{duration}, t_{now}), S, t_{now}). \quad (3.8)$$

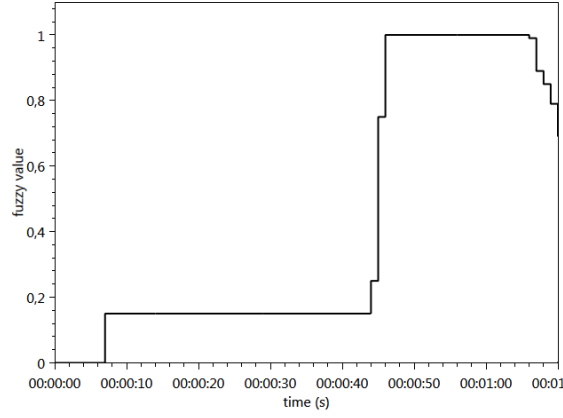


Figure 3.10: Application of the *PersWhile* predicate to the expression "the temperature is high" with a persistence of 5 seconds during the last 20 seconds.

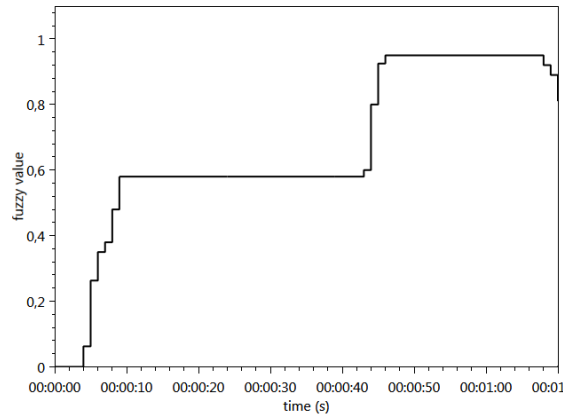


Figure 3.11: Application of the *PersWhileTol* predicate to the expression "the temperature is high" with a ratio computed on 5 seconds during the last 20 seconds.

Figures 3.10 and 3.11 show respectively these two last operators applied to the proposition "the temperature is high" for a duration of 5 seconds during the last 20 seconds. Even if the tolerant *PersWhileTol* predicate has higher values than the strict one (*PersWhile*), both of them reach the maximum value of 1 only when high temperatures are persistent, with a small delay.

Precedence

The precedence is the fact that a phenomenon occurred before a second one. This is thus a binary predicate that takes two fuzzy expressions as operands.

As a first attempt to formalize a fuzzy precedence operator from the base operators, we define the predicate " E_1 started before E_2 ".

Definition 3.12 (*Started before predicate*)

Let E_1 and E_2 be fuzzy expressions and S a scope. The *StartedBefore* predicate, as in " E_1 started before E_2 ", is defined by:

$$\text{StartedBefore}(E_1, E_2, S, t_{\text{now}}) = \text{Occ}(E_2 \wedge \text{Occ}(E_1 \wedge \neg E_2, S, t_{\text{now}}), S^*, t_{\text{now}}). \quad (3.9)$$

The deepest *Occ* predicate in equation 3.9 indicates whether in the scope there exists a

moment when the phenomenon E_1 was observed while the phenomenon E_2 was not. The precedence is true if this Occ is true and if E_2 is observed. The other Occ predicate in equation 3.9 looks for such a moment in the scope.

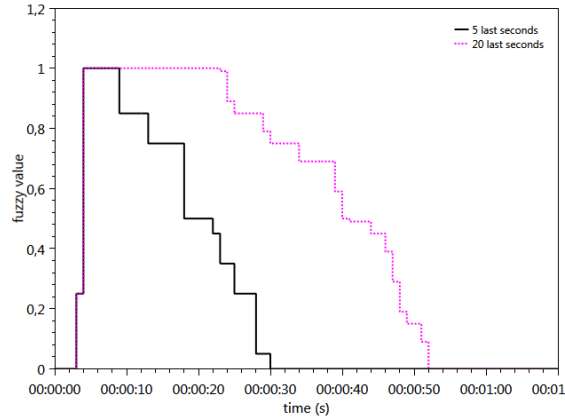


Figure 3.12: Evaluation of the expression "the rainfall is high has started before that the temperature is high" on two different scopes.

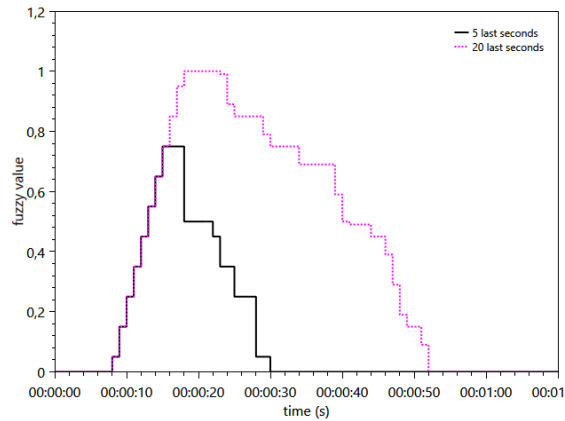


Figure 3.13: Evaluation of the expression "the rainfall is high before the temperature is high" on two different scopes.

We can define a stronger precedence relation with a *Before* predicate. The semantic of the *Before* predicate has been defined in (Allen, 1983) and has been used in many papers: to Allen, " E_1 precedes E_2 " is true if E_1 has started and has stopped before E_2 began. The *StartedBefore* predicate above does not match the semantics of Allen's precedence. To define the value of a " E_1 before E_2 " relation, E_1 must have started before E_2 , and E_1 must not be observed anymore while E_2 must have been observed.

Definition 3.13 (*Before predicate*)

Let E_1 and E_2 be fuzzy expressions and S a scope. The *Before* predicate is defined by:

$$\begin{aligned}
 Before(E_1, E_2, S, t_{now}) &= StartedBefore(E_1, E_2, S^*, t_{now}) \\
 &\quad \wedge \neg E_1(t_{now}) \\
 &\quad \wedge Occ(E_2, S, t_{now}).
 \end{aligned}
 \tag{3.10}$$

Figures 3.12 and 3.13 show respectively the application of the *StartBefore* and the *Before* predicates to the two expressions E_1 "rainfall is high" and E_2 "temperature is high". In the case of the *StartBefore* predicate, the values are high at the beginning and decrease regularly after the expression "the rainfall is high" is not true anymore. Considering a larger scope enables to maintain the activation of the precedence relation longer. Nesting the *Occ* operator prolongs the scope duration. Figure 3.13 shows that the value of operator is different from 0 until E_1 is not observed anymore in the scope.

3.2.6 Relation between *Ratio* and *Pers*

The *Ratio* operator reaches its maximum value 1 when its operand equals 1 throughout the whole scope S . Hence, it can be interpreted as a persistence of its operand. However, the difference with the *Pers* predicate is for the other values. The first difference is that *Ratio* reaches 0 only if the operand equals 0 throughout the whole scope. Then, for the intermediary values, *Ratio* is just more tolerant: its value will not be null even if the operand equals 0 only for a short while.

Thus, if a predicate uses the *Ratio* operator, it is possible to harden its behavior by replacing *Ratio* by the *Pers* predicate.

In particular, this changes the behavior of the *Increases* predicate (resp. *Decreases* predicate) that will activate only if the input is always increasing (resp. decreasing).

3.3 Spatial relations

Fuzzy logic has been successfully used in various crisis management systems. In such systems, the geographical aspect is usually very important and relies on Geographical Information Systems.

Laurence Boudet, during her postdoctoral fellowship, brought spatial reasoning to ExpressIF[®]. In particular, she adapted Fuzzy Mathematical Morphology (Bloch, 1999b) to geographic information. Chapter 2, section 2.2.5 introduces the Fuzzy Mathematical Morphology for the definition of spatial relations. During the Respondrone project, we co-supervised Clément Iphar to define spatial relations on Digital Terrain Model (DTM). DTMs are topographic models that can be manipulated by computer programs. They are usually defined as a Triangular Irregular Network.

In this section, we use the term relation instead of predicate. Indeed, in the other sections, predicates have a degree of fulfillment whereas spatial relations produce a fuzzy landscape.

3.3.1 *InTheDirection* relation

Following the definition of fuzzy landscape for cardinal directions defined by (Hudelot, Atif, and Bloch, 2008), we extended this notion to any angle $G \in [0, 2\pi[$ to define the spatial relation "in the direction G ".

Let ν_G^P be the structuring element of this relation. We chose simple parameters to customize it: θ_1 , the lower cut angle, and θ_2 , the upper cut angle. Considering each point of the space \mathbb{R}^3 , with East, North and elevation coordinates, the problem is reduced to East and North coordinates, since the membership value of a point does only depend on the azimuth angle and does not depend on the elevation angle.

Let us consider the point X of coordinates (E_X, N_X, z_X) , and the origin of the structuring element O of coordinates $(0, 0, 0)$. The bearing of vector \overrightarrow{OX} with respect to the North (i.e. a positive angle measured clockwise with respect to the North) is computed as $\omega_{OX} = \text{atan2}(E_X, N_X)$.

Definition 3.14 (InTheDirection relation)

The *InTheDirection* relation is based on the fuzzy morphomathematics framework. Its resulting fuzzy landscape is given by the application of the fuzzy dilation to a fuzzy object of interest B (whose membership function is μ_B). Let the result of this dilation of B by the structuring element ν_G^P at each point X be $\mu_G^P(B)(X)$:

$$\forall X \in \mathbb{R}^3,$$

$$\mu_G^P(B)(X) = \mu_B(X) \vee \sup_{b \in B} (\mu_B(b) \wedge \nu_G^P(b)(X)) \quad (3.11)$$

with the structuring element ν_G^P defined as: $\forall X \in \mathbb{R}^3 \setminus O$,

$$\nu_G^P(O)(X) = \begin{cases} 1 & \text{if } \Delta_X^G \leq \theta_1 \\ \cos \frac{\Delta_X^G - \theta_1}{\theta_2 - \theta_1} & \text{if } \theta_1 < \Delta_X^G < \theta_2 \\ 0 & \text{if } \Delta_X^G \geq \theta_2. \end{cases} \quad (3.12)$$

where O is the origin of the structuring element and $\Delta_X^G \in [0, \pi]$ the angular difference between ω_{OX} and G as seen from point O .

By construction, $\theta_1 \leq \theta_2$ and it is expected that $\theta_2 \leq \frac{\pi}{2}$ to avoid getting positive membership values for points that are completely outside the range of coordinates in the desired direction. Two instances of this structuring element are shown in Figure 3.14. As usual in this manuscript, the white color represents a value of 0 while the black color represents a value of 1.

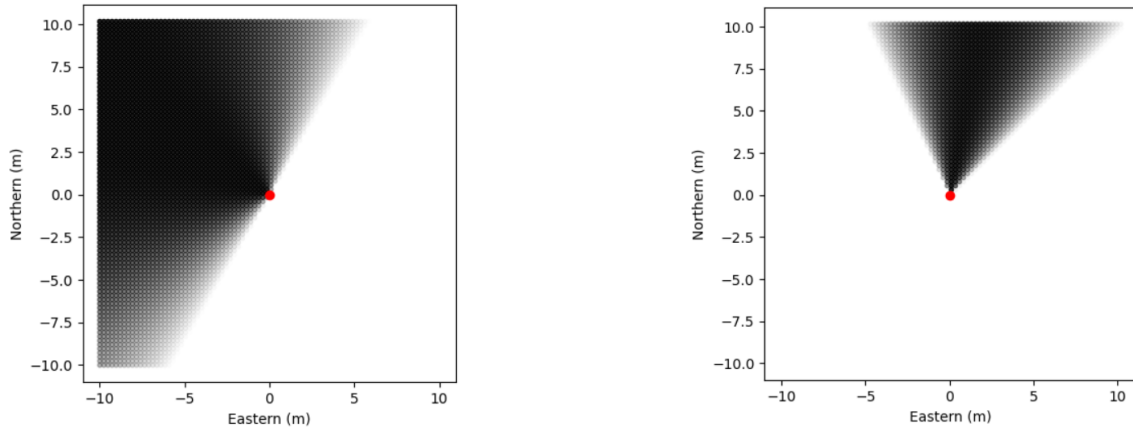


Figure 3.14: Two structuring elements ν_G^P representing the relation "in the direction G ". The red dot represents the center of the structuring element. For the left picture, $G = 300^\circ$, $\theta_1 = \frac{\pi}{12}$ and $\theta_2 = \frac{\pi}{2}$; for the right picture, $G = 10^\circ$, $\theta_1 = 0$ and $\theta_2 = \frac{\pi}{5}$.

The relation reaches its maximum for the points of the space for which the angular difference to G is inferior to the lower cut angle θ_1 . It reaches its minimum for the points of the space for which the angular difference to G is superior to θ_2 . For the other points, the value of the relation decreases as Δ_X^G increases, following a trigonometric function. Here the choice of the trigonometric function is arbitrary, other functions (linear or not, continuous or not, as long as defined at all points) could be used.

Figure 3.15 shows the dilation of objects (in red) by the two structuring elements shown in Figure 3.14.

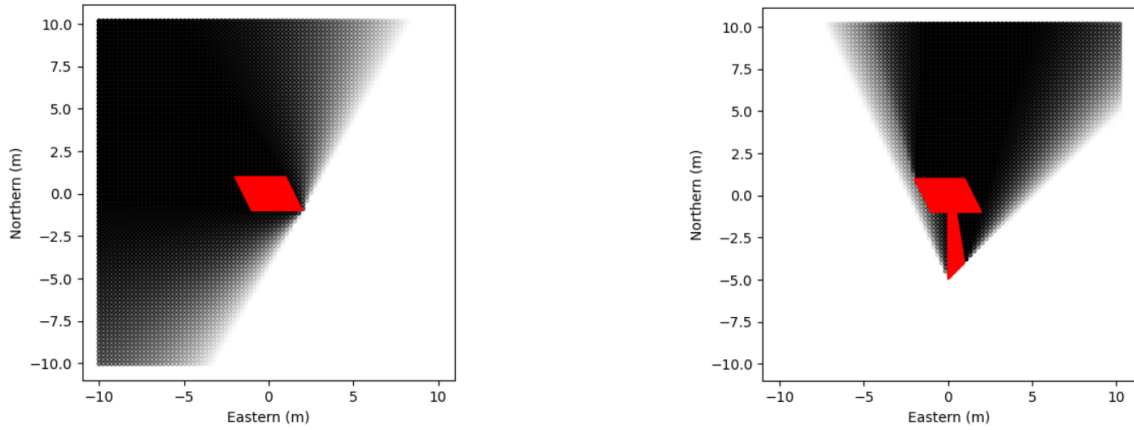


Figure 3.15: Fuzzy landscape of two objects (crisp in this example) dilated by ν_G . For left figure, $G = 300^\circ$, $\theta_1 = \frac{\pi}{12}$ and $\theta_2 = \frac{\pi}{2}$; for the right figure, $G = 10^\circ$, $\theta_1 = 0$ and $\theta_2 = \frac{\pi}{5}$.

Near relation

On top of the general consideration towards the direction, a consideration on the distance to the object can be considered, leading to the use of both *Near* relation and *InTheDirection* relation and thus creating the *NearAndInTheDirection* relation. In this respect, we defined the *Near* relation that has a structuring element denoted as $\nu_n^{P'}$, taking the set of parameters $P' = \{\delta_1, \delta_2\}$. Those parameters are δ_1 , which is the lower cut distance, or the greatest distance for which the membership score is 1 and δ_2 , which is the upper cut distance, or the smallest distance for which the membership score is 0. As one of the elementary spatial primitives, as defined in (Freeman, 1975), its definition is kept as simple as possible, to be later combined with other spatial relations.

Definition 3.15 (*Near structuring element*)

Denoting $\mathcal{D}_E(O, X)$ as the Euclidean distance between O and X , $\forall X \in \mathbb{R}^3$, $\delta_1 \geq 0$, $\delta_2 \geq \delta_1$

$$\nu_n^{P'}(O)(X) = \begin{cases} 1 & \text{if } \mathcal{D}_E(O, X) \leq \delta_1 \\ 1 - \frac{\mathcal{D}_E(O, X) - \delta_1}{\delta_2 - \delta_1} & \text{if } \delta_1 < \mathcal{D}_E(O, X) < \delta_2 \\ 0 & \text{if } \mathcal{D}_E(O, X) \geq \delta_2 \end{cases} \quad (3.13)$$

Figure 3.16 shows the membership function of $\nu_n^{P'}$ and a representation of the structuring element in the (E, N) , in the horizontal plane that is at $z = 0$. Please take into consideration that the structuring element is actually of spherical shape.

NearAndInTheDirection relation

The novelty of our approach consists in considering both spatial relations of direction and proximity in one single structuring element applied to the object of interest. To perform a fuzzy intersection between the two concepts, we use a t-norm as in (Bloch and Maitre, 1995). For the sake of simplicity and without loss of generality, the only t-norm that will be used in all illustrations of this section is the product.

Following the notations of Eq.3.12 and Eq.3.13, the membership value of the dilation of B by ν_G^P and $\nu_n^{P'}$ at each point x is denoted $\mu_{NG}^{P, P'}(B)(x)$ and computed as the supremum of the membership values of the t-norm operation on ν_G^P and $\nu_n^{P'}$ when applied to each

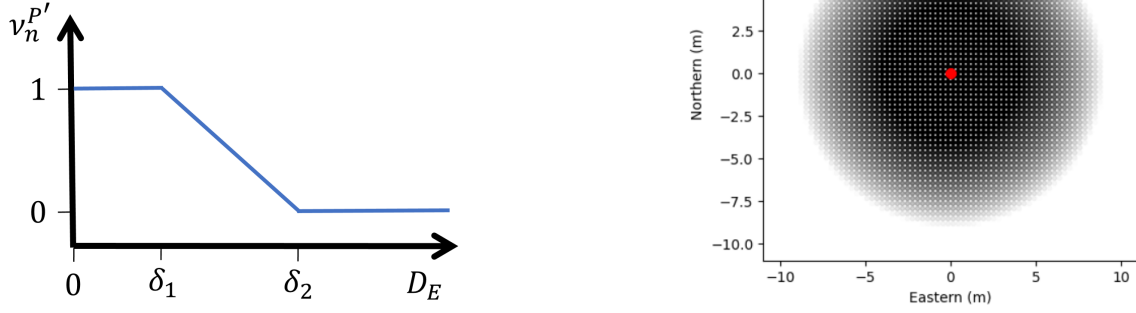


Figure 3.16: Left: membership function for the "Near" relation. Right: corresponding structuring element ν_n with $\delta_1 = 4$ and $\delta_2 = 9$.

point of the initial set.

Definition 3.16 (*NearAndInTheDirection relation*)

The fuzzy landscape resulting of the application of the *NearAndInTheDirection* relation on the reference object B is given by the membership function $\mu_{NG}^{P,P'}$ defined by: $\forall X \in \mathbb{R}^3$,

$$\mu_{NG}^{P,P'}(B)(X) = \perp(\mu_B(X), \sup_{b \in B} (\top(\mu_B(b), \nu_{NG}^{P,P'}(b)(X)))). \quad (3.14)$$

and whose structuring element ν_{NG} is computed in each point X of \mathbb{R}^3 from any point $b \in B$ as

$$\nu_{NG}^{P,P'}(b)(X) = \top(\nu_G^P(b)(X), \nu_n^{P'}(b)(X)). \quad (3.15)$$

Figure 3.17 shows the dilation of the same objects with the same parameters of Figure 3.15 but with the structuring element ν_{NG} instead of ν_G

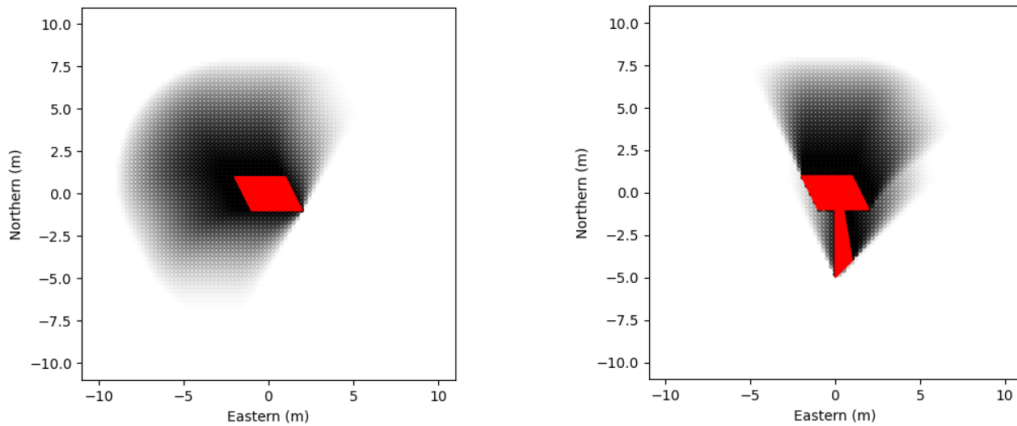


Figure 3.17: The same objects of Figure 3.15 dilated by ν_{NG} with the same sets of parameters, adding $\delta_1 = 2$ and $\delta_2 = 7$ for both pictures, considered in the plane of interest ($z = 0$).

3.3.2 Definition of the ground level as a fuzzy landscape

In this section, we are interested in creating a fuzzy landscape for the determination of the ground level, i.e. the area of space that is just above the ground. This landscape covers the whole area of interest, and rather than considering the geometrical distance between the surface (represented by a collection of polygons) and the point of interest, we considered the vertical distance between the point and the surface, considering the only direction that matters in terms of height: the z-axis.

Let us consider a point X of coordinates (E_X, N_X, z_X) in \mathbb{R}^3 . Let us compute z_0 , the altitude according to the DTM of interest. Let us define $\Delta z = z_X - z_0$ as the height above the ground at the point of interest. Figure 3.18 shows such membership function, as defined by Eq.3.16.

$$\forall X \in \mathbb{R}^3, \tau_1 \geq 0, \tau_2 \geq \tau_1,$$

$$\mu(\Delta z) = \begin{cases} 0 & \text{if } \Delta z < 0 \\ 1 & \text{if } 0 \leq \Delta z \leq \tau_1 \\ 1 - \frac{\Delta z - \tau_1}{\tau_2 - \tau_1} & \text{if } \tau_1 < \Delta z < \tau_2 \\ 0 & \text{if } \Delta z \geq \tau_2 \end{cases} \quad (3.16)$$

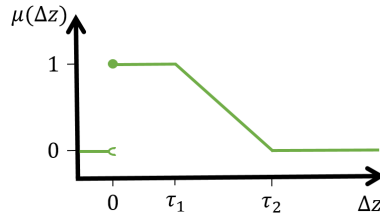


Figure 3.18: Membership function "at ground level" with respect of the computed value of Δ_z for any point X of \mathbb{R}^3 .

We then applied the simple membership function shown in Figure 3.18 to the whole surface of the DTM, by applying the $\mu(\Delta z)$ function to all points of the terrain within the boundary box of interest, according to the local height at each point. The membership function of the resulting 3D fuzzy set is denoted μ_T . An illustration of the terrain and of the resulting fuzzy set computed with the set of parameters $\tau_1 = 1 \text{ m}$ and $\tau_2 = 3 \text{ m}$ is shown in Figure 3.19.

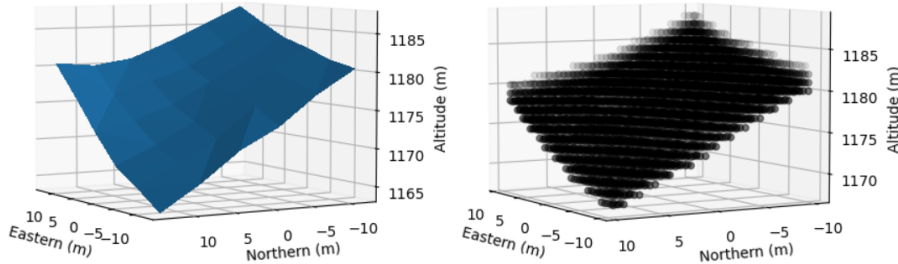


Figure 3.19: Left: 20m \times 20m excerpt of the DTM. Right: Fuzzy set for the spatial relation "At Ground Level" on the same DTM area. Centered on $E = 1201690$, $N = 6147769$, for the sake of the example.

3.4 Online spatio-temporal predicates

With the addition of temporal and spatial predicates in ExpressIF[®], it was natural to formalize spatio-temporal ones. I supervised the post-doctoral fellowship of Jean-Marie Le

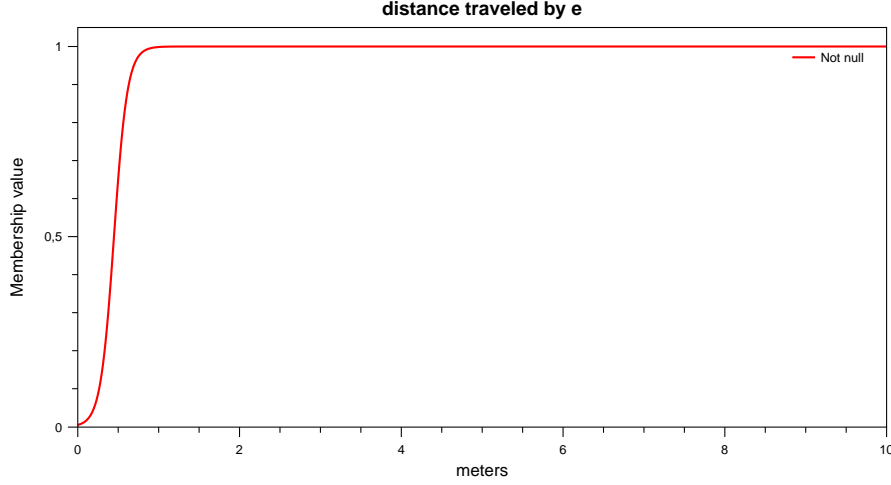


Figure 3.20: Example of a membership function used for estimating $P_{>0}$ ="the distance traveled by e since the previous position is not null". The distance is expressed in meters (m).

Yaouanc in this aim. In this work, we considered objects that are geolocalized, for instance by a typical GPS sensor. In other words, we considered that the inputs of our system are streams of point locations from localization sensors, which give no more information than the timestamped coordinates (i.e. no information about the nature, the shape or the condition of the entity).

3.4.1 Base predicates

We defined the *IsMoving* predicate as a base predicate that indicates if the entity of interest is moving, i.e. the distance between two positions is strictly greater than 0, throughout a fuzzy scope.

Definition 3.17 (*IsMoving predicate*)

The predicate *IsMoving*, applied on an entity of interest e , throughout a fuzzy scope S at time t_{now} is defined by:

$$IsMoving(e, S, t_{now}) = P_{>0}(t) \vee Ratio(P_{>0}, S^*, t_{now}) \quad (3.17)$$

where $P_{>0}$ is the fuzzy proposition "the distance traveled by e since the previous position is not null". We voluntarily omit $P_{>0}$ in the operator parameter list for the sake of readability.

The nested proposition $P_{>0}$ handles the spatial uncertainty, specifically because of the inaccuracy of GPS sensors. Figure 3.20 presents an example of a membership function suitable for the proposition $P_{>0}$. The *IsMoving* predicate is thus describing the fact that the entity e is moving without any direction or speed considerations. The disjunction in the formula is important because it handles the temporal uncertainty of this operator: either e has traveled a certain distance between the two last known positions, or it happened in the recent past. For instance, if e is a walking human, if he drops its keys and stops to pick them up, regarding the application we may want to consider the overall behavior, i.e. that he is walking. Another example with a car: we may want to characterize the overall trajectory without taking into account the stops at traffic lights.

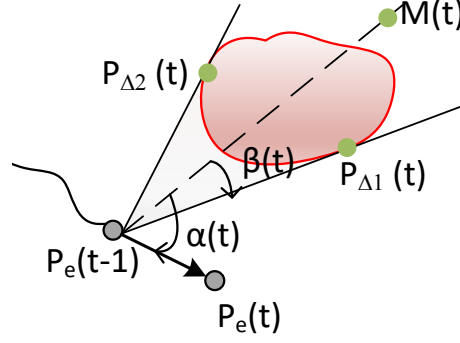


Figure 3.21: Elements of comprehension for the *IsGoingCloseTo* predicate.

3.4.2 Trajectory predicates

To describe the next predicates, let us introduce two notations:

- \bar{o} is the exterior of a region o ;
- $\in(e, o, t)$ is a geometrical predicate that states if e belongs to a certain region o at time t , either crisp or fuzzy: in the latter case, it supports *IsMoving* operator in the handling of the spatial uncertainty.

Definition 3.18 (*IsGoingCloseTo* predicate)

The *IsGoingCloseTo* predicate indicates if an entity e is approaching a region o throughout a scope S :

$$\begin{aligned}
 \text{IsGoingCloseTo}(e, o, S, t_{\text{now}}) &= \text{Ratio}(\in(e, \bar{o}, t_{\text{now}}), S, t_{\text{now}}) \wedge \text{IsMoving}(e, S, t_{\text{now}}) \\
 &\quad \wedge ((E_{\text{dir}}(t_{\text{now}}) \wedge E_{\text{close}}(t_{\text{now}})) \\
 &\quad \vee \text{Ratio}(E_{\text{dir}} \wedge E_{\text{close}}, S^*, t_{\text{now}})).
 \end{aligned} \tag{3.18}$$

where E_{dir} is measuring how much e is directing towards o and E_{close} indicates if e is close enough to o .

Figure 3.21 shows some clues of the computation of E_{dir} . At time t , we consider that the optimal direction towards o is given by $\overrightarrow{P_e(t-1)M(t)}$ and that the acceptable directions are inside the cone $P_e(t-1)P_{\Delta 1}(t)P_{\Delta 2}(t)$. We thus use angles $\alpha(t)$ and $\beta(t)$ and a fuzzy set to quantify the adequacy of the direction. $E_{\text{dir}}(t)$ is defined by the expression " $D_{to}(t)$ is close to 1" where

$$D_{to}(t) = \min(\cos(\max(0, \alpha(t) - \beta(t))), 0). \tag{3.19}$$

Exploiting the same geometrical formalization of the direction, we defined the opposite predicate *IsGoingAway*. Let $E_{\text{op.dir}}$ be the fuzzy proposition "the entity e is moving in the opposite direction of the region o ", which can be expressed by $E_{\text{op.dir}}(t) = "D_{op}(t)$ is close to -1 " where

$$D_{op}(t) = \min(\cos(\alpha(t)), 0). \tag{3.20}$$

Definition 3.19 (*IsGoingAway* predicate)

The *IsGoingAway* predicate indicates if an entity e is going away from a region o throughout the scope S and is defined by:

$$\begin{aligned}
IsGoingAway(e, o, S, t_{now}) &= Ratio(\in(e, ^-o, t_{now}), S, t_{now}) \\
&\wedge IsMoving(e, S, t_{now}) \\
&\wedge (E_{op.dir}(t_{now}) \vee Ratio(E_{op.dir}, S^*, t_{now})).
\end{aligned}
\tag{3.21}$$

This definition of the *IsGoingAway* predicate is here intended to a local use. Actually, it did not use a proximity predicate: in this case, an object moving in the opposite direction of the region, but located very far from it, will be considered going away. For a general use, a fuzzy proposition stating the proximity of the entity to the region must be added to the equation 3.21, as we did in section 3.3.1.

3.4.3 Entrance and exit predicates

Let us consider the fact that an entity e is entering a closed geometrical surface o with membership function μ_o indicating how much a point belongs to o . The object o can be crisp or fuzzy. The formulation of such an operator is quite straightforward given the existing operators: entering a closed geometrical surface means being outside this area before being inside this area.

Definition 3.20 (*IsEntering* predicate)

The *IsEntering* predicate indicates if an entity e is entering a region o throughout the scope S and can be defined by:

$$\begin{aligned}
IsEntering(e, o, S, t_{now}) &= IsMoving(e, S, t_{now}) \\
&\wedge Before(\in(e, ^-o, t_{now}), \in(e, o, t_{now}), S, t_{now}).
\end{aligned}
\tag{3.22}$$

In practice, the scope S must be a little greater than the sampling rate of the position of e : if S is too big, *IsEntering* will trigger only once during the scope S because of *Before* operator: so cases of a reentrance will not be detected.

We can derive from *IsEntering* the *IsExiting* operator.

Definition 3.21 (*IsExiting* predicate)

The *IsExiting* predicate indicates if an entity e is exiting a region o throughout the scope S and can be defined by:

$$\begin{aligned}
IsExiting(e, o, S, t_{now}) &= IsMoving(e, S, t_{now}) \\
&\wedge Before(\in(e, o, t_{now}), \in(e, ^-o, t_{now}), S, t_{now}).
\end{aligned}
\tag{3.23}$$

3.4.4 Compartmentalization predicates

This type of predicates monitors the entity e and states about its behavior in two ways: either it keeps moving in a closed area, or it is following a path that it has been asked to follow.

On the one hand, the first operator is straightforward to define. It states if e keeps moving while staying in a given closed area o .

Definition 3.22 (*IsMovingInside predicate*)

We defined the *IsMovingInside* predicate with the following equation:

$$IsMovingInside(e, o, S, t_{now}) = E_{MovingIn}(t_{now}) \wedge Ratio(E_{MovingIn}, S^*, t_{now}) \quad (3.24)$$

where the fuzzy expression $E_{MovingIn}$ evaluated at time t states if e is moving inside o at that moment. It can be defined by:

$$E_{MovingIn}(t) = P_{>0}(t) \wedge \in(e, o, t). \quad (3.25)$$

The disjunction in equation 3.24 handles the spatial uncertainty: for instance, if e is out of o for a brief moment, the ratio will smooth this fact. As for *IsExiting* and *IsEntering* operators, the inclusion relationship $\in(e, o, t_{now})$ between e and o can be either crisp or fuzzy.

On the other hand, to monitor if e is following a path, we defined a path Π , also called an itinerary, as an oriented polyline. Thus, to follow a path, the entity e must move in the same direction of Π and must be close enough to Π while moving (figure 3.22).

Let $P_e(t)$ be the position of e at time t . Let $\overline{QQ'}$ the line segment of Π which $P_e(t)$ has been projected onto. Thus, the distance $d(e, \overline{QQ'})$ between e and $\overline{QQ'}$ must be kept low. Let E_{close} be the fuzzy proposition " $d(e, \overline{QQ'})$ is low".

Moreover, we considered the orientation of e regarding Π . We could use exogenous orientation information (e.g. a magnetometer), but it is not a good characterization of the trajectory orientation. For instance, imagine a human moving sideways: his own orientation is thus perpendicular to the trajectory. We use the orientation of $\overrightarrow{P_e(t-1)P_e(t_{now})}$ instead, which is thus deduced from the last move; consequently, we need two positions to start characterizing the behavior. If e follows exactly Π , $\overrightarrow{QQ'}$ and $\overrightarrow{P_e(t)P_e(t)}$ are collinear and have the same direction. So if we note $\alpha'(t)$ the angle between the two vectors at time t , $\cos(\alpha'(t)) = 1$. Let E'_{dir} be the fuzzy proposition " e is moving in the good direction". For instance, the value of this proposition can be evaluated by $\max(\cos(\alpha'(t)), 0)$: the more $\max(\cos(\alpha'(t)), 0)$ tends to 1, the more e is moving in the right direction. Obviously, $\max(\cos(\alpha'(t)), 0)$ equals 0 for an opposite direction.

Definition 3.23 (*IsFollowingAPath predicate*)

The *IsFollowingAPath* predicate indicates if an entity e is following the path Π throughout the scope S and can be defined by:

$$\begin{aligned} IsFollowingAPath(e, \Pi, S, t_{now}) &= IsMoving(e, S, t_{now}) \\ &\wedge (E_{close}(t_{now}) \vee Ratio(E_{close}, S^*, t_{now})) \\ &\wedge (E'_{dir}(t_{now}) \vee Ratio(E'_{dir}, S^*, t_{now})). \end{aligned} \quad (3.26)$$

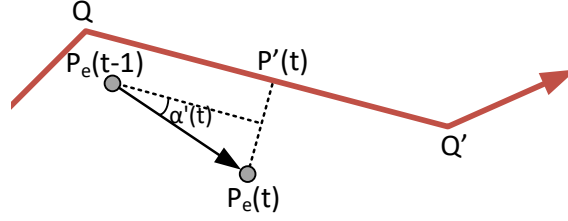


Figure 3.22: Illustration of the different elements of the *IsFollowingAPath* operator equation.

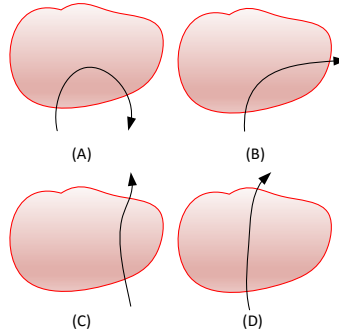


Figure 3.23: Different examples of trajectories which cross the region, adapted from (Vanegas Orozco, 2011).

In the previous definition, we voluntarily omitted E_{close} and E'_{dir} in the parameter list for the sake of readability. Both expressions are important parameters because they allow customizing both the temporal and the spatial uncertainty of the operator regarding the application.

3.4.5 Crossing predicate

In this section, we present a predicate to assess if e is crossing a closed area o . This predicate is not easy to define and several attempts have been done in the past. In (Vanegas Orozco, 2011), the author studied the different meanings of the term "crossing" and asked a pool of 32 persons to choose among 8 figures which ones define the best "crossing". The conclusion is that "crossing" the points on which the entity enters and goes out of the region are located on "opposite" sides. Figure 3.23 shows different itineraries that cross the region. The examples are given in increasing order of value for an intuitive definition of "crossing". The example (A) is not a case of crossing because the entity enters and leaves on the same side of the region. In (B), the entity has crossed the region but the sides are not opposed. The case (C) and (D) are really "crossing" examples but the path inside the region is shorter in (C) than in (D), so we expect a lower value in (C) than in (D).

We thus chose to introduce a criterion to assess how much e is crossing o . During the crossing of o , the trajectory of e is splitting o into two closed geometries $\mathcal{O}_1(t)$ and $\mathcal{O}_2(t)$, whose areas are respectively $\mathcal{A}_1(t)$ and $\mathcal{A}_2(t)$ (figure 3.24). Our criterion establishes that the value of "e is crossing o" takes its maximum value when it splits o into $\mathcal{O}_1(t)$ and $\mathcal{O}_2(t)$ such as $\mathcal{A}_1(t) = \mathcal{A}_2(t)$. Of course, it is not sufficient: e has to be moving inside o .

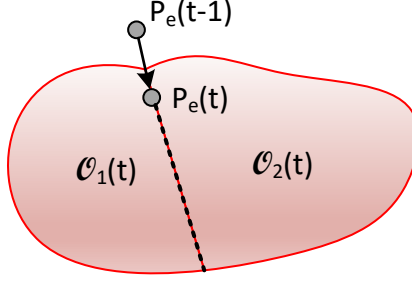


Figure 3.24: Illustration of the different elements of the *IsCrossing* operator equation.

Definition 3.24 (*IsCrossing predicate*)

The *IsCrossing* predicate indicates if an entity e is going through a region o throughout the scope S and can be defined by:

$$\begin{aligned}
 IsCrossing(e, o, S, t_{now}) &= IsMoving(e, S, t_{now}) \wedge \in(e, o, t_{now}) \\
 &\quad \wedge (Criterion(t_{now}) \vee Ratio(Criterion(t), S^*, t_{now})).
 \end{aligned}
 \tag{3.27}$$

where $Criterion(t)$ is a function whose values are in $[0,1]$ indicating if the two areas $\mathcal{A}_1(t)$ and $\mathcal{A}_2(t)$ are equal. It has to equal 1 when the two areas are equal.

As usual, we handle the uncertainty about the general trajectory with the *Ratio* operator.

In our implementation, $Criterion(t)$ is based on the entropy:

$$Criterion(t) = -\frac{\mathcal{A}_1(t)}{\mathcal{A}} \log_2 \frac{\mathcal{A}_1(t)}{\mathcal{A}} - \frac{\mathcal{A}_2(t)}{\mathcal{A}} \log_2 \frac{\mathcal{A}_2(t)}{\mathcal{A}}
 \tag{3.28}$$

where \mathcal{A} is the area of o .

To compute $\mathcal{A}_1(t)$ and $\mathcal{A}_2(t)$ when e is inside o but not on the boundary, we have to extrapolate its trajectory (see the dotted line in figure 3.24): although there are a lot of extrapolation methods, we chose to simply extend the last line segment of the trajectory until it intersects the boundary of o . At each new position, the frontier between $\mathcal{O}_1(t)$ and $\mathcal{O}_2(t)$ is re-evaluated with the past positions and the extrapolation of the itinerary.

3.5 Knowledge about the predicates

The composition paradigm that allows defining new predicates from existing ones emphasizes there exist relationships between those predicates. The intuition is that those relationships can help bounding the values of predicates.

Our first attempt to show off such relations has been made for the spatio-temporal predicates and gave the subsumption lattice as shown in figure 3.25. In this figure, the symbols \top and \perp are borrowed from description logics and represent the universal and the empty concepts. Figure 3.25 only shows the relations between a mobile entity and a region (i.e. it does not include the *IsFollowingAPath*). For instance, as we characterize moves, all the relations are more specific than *IsMoving*, and *IsGoingThrough* is more specific than a scenario consisting in the succession of *IsEntering*, *IsMovingInside* and *IsExiting*. In other words, if the entity e is not moving, we can infer that it is not going along, not going close to, etc.

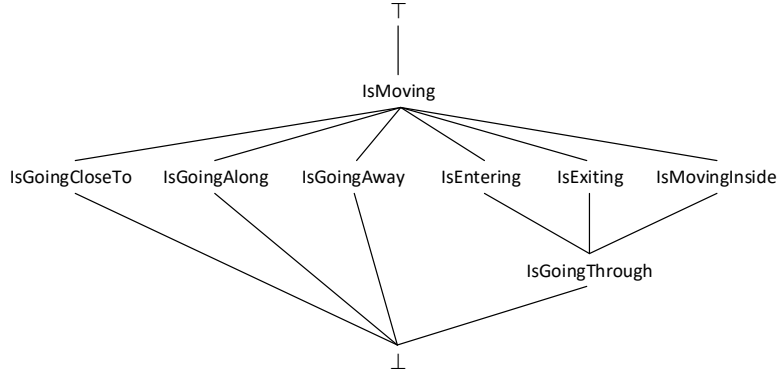


Figure 3.25: Subsumption lattice for the spatio-temporal relations (Poli, Boudet, and Le Yaouanc, 2018).

Within Regis Pierrard’s PhD, we went further in the consideration of the relationships between predicates in order to improve the performances of predicate learning (see chapter 4, section 4.3). We identified three main kinds of links between two predicates:

- Dependency: a predicate P_1 is dependent on a predicate P_2 if the evaluation of P_2 is necessary to the evaluation of P_1 .
- Symmetry: a predicate P is symmetric if it has the same value for any permutation of its operands.
- Implication: we consider different types of implications but the intuition behind this link is that it allows propagating the values from one predicate to the other. For instance, we can express that if a predicate P_1 equals 1 then P_2 should also equal 1.

More details can be found in (Pierrard, 2020). We thus consider a labeled directed graph to represent this knowledge about the predicates. In particular, we consider the six types of links that are shown in table 3.1.

Link	Notation	Corresponding edge
\mathcal{P}_2 depends on \mathcal{P}_1	d	$\mathcal{P}_2 \xrightarrow{d} \mathcal{P}_1$
\mathcal{P}_1 is symmetrical	c	$\mathcal{P}_1 \xrightarrow{c} \mathcal{P}_1$
$\mathcal{P}_1 \Rightarrow \mathcal{P}_2$	i	$\mathcal{P}_1 \xrightarrow{i} \mathcal{P}_2$
$\mathcal{P}_1 \Rightarrow \overline{\mathcal{P}_2}$	e	$\mathcal{P}_1 \xrightarrow{e} \mathcal{P}_2$
$\overline{\mathcal{P}_1} \Rightarrow \mathcal{P}_2$	ni	$\mathcal{P}_1 \xrightarrow{ni} \mathcal{P}_2$
$\overline{\mathcal{P}_1} \Rightarrow \overline{\mathcal{P}_2}$	ne	$\mathcal{P}_1 \xrightarrow{ne} \mathcal{P}_2$

Table 3.1: The different kinds of link between relations and their notation in the graph representation. The third column specifies how the corresponding edge is represented in a graph. \mathcal{P}_1 and \mathcal{P}_2 are two p -ary fuzzy predicates (Pierrard, 2020).

Let us take as example nine spatial predicates (also called spatial relations). Four of them are the directional relations described by Bloch (Bloch, 1999b): *to the left of*, *above*, *to the right of* and *below*. Four other relations are also directional. They express the same directions (left, right, above, below) but their fuzzy landscapes cover a smaller area of the image to express the following relations: *completely to the left of*, *completely to the right of*, *completely above* and *completely below*. The last relation is the symmetry measure that is presented in (Colliot, 2003). The goal of this operator is to assess if two objects are symmetrical. The knowledge graph about these predicates is shown in figure 3.26.

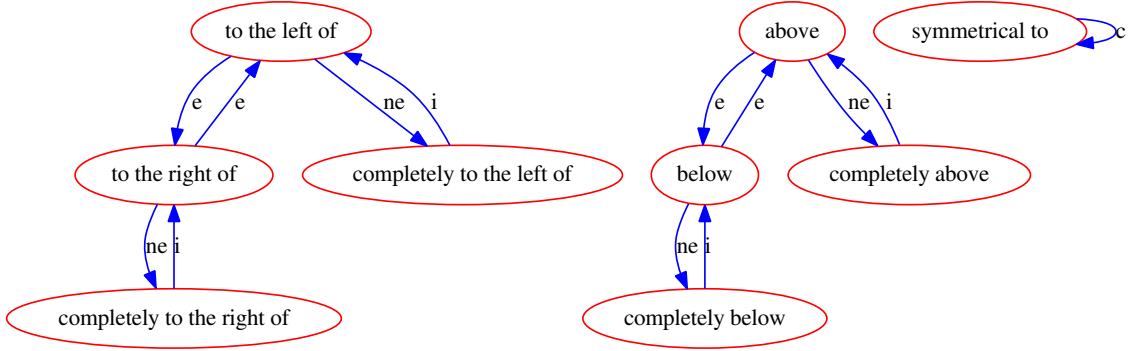


Figure 3.26: Graph representing the logical links between the nine spatial predicates considered in the example (Pierrard, 2020).

3.6 Applications

3.6.1 Crisis management

ExpressIF[®] is a key tool for crisis management. It has been involved in many collaborative and industrial projects in that field (e.g. Descartes, Respodrone to name a few). Using rules allows basing decision on the aggregation of heterogeneous data (e.g. weather forecast, geographical data, sensors data, etc.). In addition, before the crisis, a variety of procedures and expertise are available. In particular, the spatial, temporal and spatio-temporal predicates are involved to model first-responders and specialists knowledge.

Those predicates can also be used to make some assumptions about the damages caused by the crisis. For instance, in the work of Clément Iphar, we were interested in predicting run-off areas using commonsense reasoning instead of a modeling in fluid mechanics, using the spatial predicates presented in section 3.3. Indeed the run-off areas are at the ground level and in the direction of the greatest dip.

We describe here an application around a spill of liquid from a building. The spatial extent of the building, denoted B , is known, as well as the local terrain. The computation process takes three steps:

1. the computation of all runoff tracks from all the points of the object, defined as a fuzzy set whose membership function is denoted μ_R ;
2. the dilation of μ_R by ν_{NG} , the structuring element considered at the point of interest;
3. the consideration of the ground surface characterized by the membership function μ_T as defined in section 3.3.2.

In this section, all the figures are shown with the following set of parameters: $P = [\frac{\pi}{12}, \frac{\pi}{4}]$, $P' = [5, 20]$, and when necessary $\tau_1 = 1$ and $\tau_2 = 3$.

When B is a fuzzy object, the collection of runoff tracks is a fuzzy set whose membership value μ_R for each point of the runoff being determined from the membership value of the point of B from which it originated. We define the resulting fuzzy set, characterized by μ_R , as the fuzzy union of all runoff tracks from all points of the object B . We compute this set in several steps. First, we denote all points originating from the b^{th} element of B as the fuzzy set characterized by μ_{R_b} :

$$\mu_{R_b} = \{(X, \mu_B(b)) | X \in R_{E_b, N_b}\} \quad (3.29)$$

taking $\mu_B(b)$ as membership value of B at point b and where R_{E_b, N_b} is the set of points computed by Procedure 1 given that E_b and N_b are the coordinates of point b .

Then, we compute μ_R with the following equation:

$$\mu_R = \bigcup_{b \in B} \mu_{R_b} \quad (3.30)$$

where \bigcup is the fuzzy union of two fuzzy sets such that $\forall b, b' \in B, b \neq b'$,

$$\mu_{R_b} \bigcup \mu_{R_{b'}} = \{(X, \max(\mu_{R_b}(X), \mu_{R_{b'}}(X))) | X \in R_{E_b, N_b} \bigcup R_{E_{b'}, N_{b'}}\}. \quad (3.31)$$

According to Eq. 3.14, the raw runoff fuzzy set is dilated by ν_{NG} and the resulting set is denoted μ_B^{NG} , and computed $\forall X \in \mathbb{R}^3$ as

$$\mu_B^{NG}(X) = \perp(\mu_R(X), \sup_{b \in B} (\top(\mu_R(b), \nu_{NG}(b)(X)))). \quad (3.32)$$

The last step consists in performing the intersection of μ_B^{NG} and μ_T . The resulting fuzzy set, the fuzzy runoff area, denoted μ_B^* , is computed as, $\forall X \in \mathbb{R}^3$,

$$\mu_B^*(X) = \top(\mu_B^{NG}(X), \mu_T(X)). \quad (3.33)$$

Figure 3.27 shows the outcome of the computation for the same runoff in both 2D and 3D visual representation.

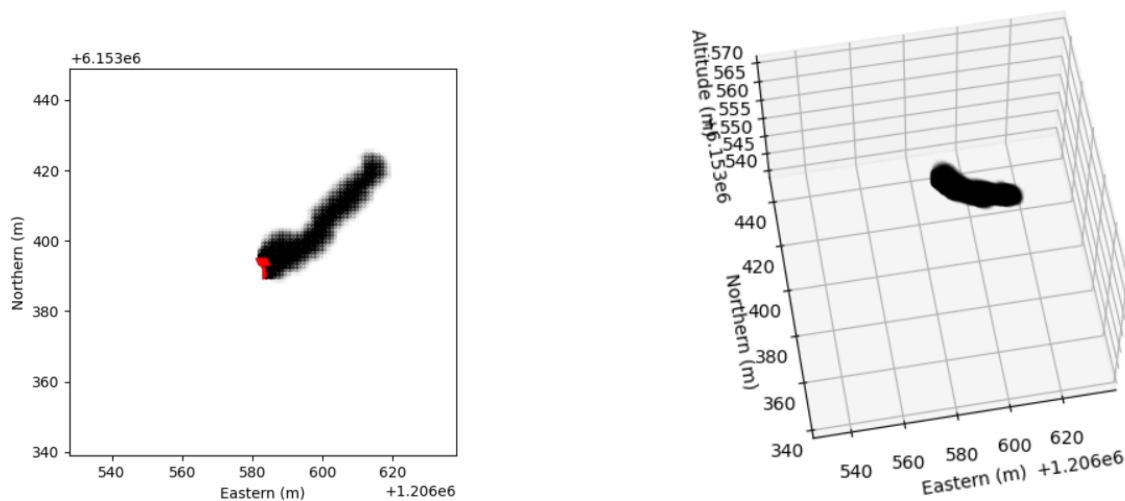


Figure 3.27: Representations of the runoff area. Left: 2D from above. Right: 3D.

This method allows highlighting the cases in which various parts of the initial set (the building) will lead to distinct runoff areas, due to the local topography. While Figure 3.27 shows a rather homogeneous runoff direction, Figure 3.28 shows two different runoff areas that are constituted by a plurality of branches.

Finally, let us show a result on a real building, Corte castle, located in Corte, France. It was supposed to be the place of the final demonstration of the Respodrone project (the final demonstration took place in Spain, due to the pandemic). We extracted its geometry using publicly available data from Open Street Map¹. This castle, located at the top of a rocky peak, displays complex topographic surroundings, and the generation of a runoff map would be advantageous to first responders. Figure 3.29 shows a picture of the castle, to underline its topographic features.

The outcome of the computation is shown in Figure 3.30, showing for instance the discharge ways of water in the event of intense rainfalls.

¹<https://www.openstreetmap.org/>

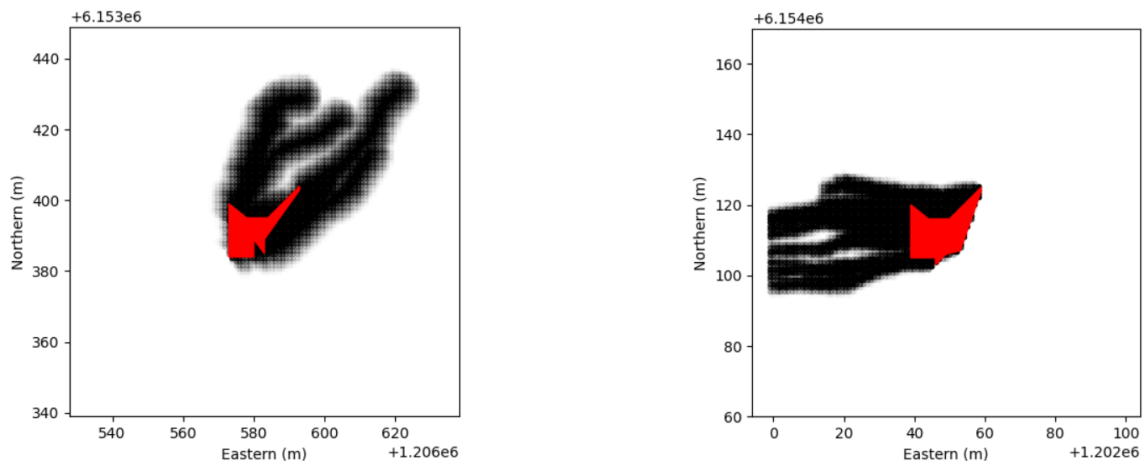


Figure 3.28: Examples of runoff from crisp polygons, with various shapes and from different points on the DTM.



Figure 3.29: View of the Corte castle (from South).

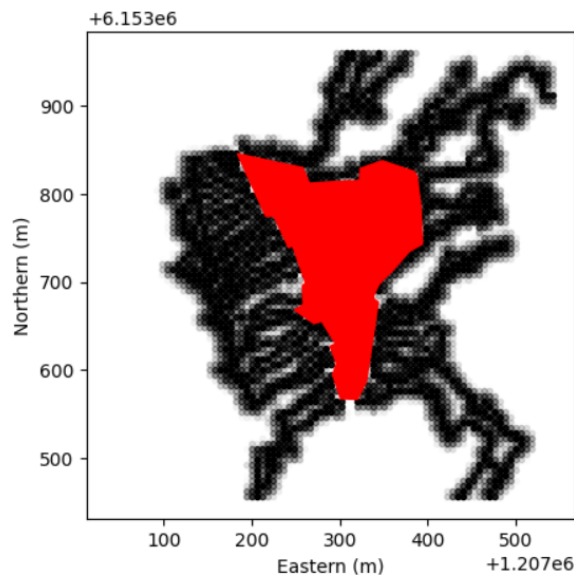


Figure 3.30: Runoff area from the Corte castle.

3.6.2 Water treatment plant management

The temporal predicates are useful when it comes to manage a complex system. The occurrence allows detecting an event even if it is low and the persistence allows to wait for an event to last a given amount of time before making some decisions (potentially decreasing the number of false alarms). We interviewed some experts of a water treatment plant near Abu Dhabi to help with maintenance. The rule base, that cannot be public, is able to suggest additional analyzes or corrective actions.

The Abu Dhabi plant is one of the most modern among the plants managed by the company we worked with. Many sensors are used to measure the pH, the concentration of some compounds, the temperatures, etc. Nevertheless, some measurement cannot be performed by sensors: for instance, the color of water must be checked by human workers. Some analysis of the water are also based on water samples that are sent to a laboratory.

The rule base has thus inputs that are directly sensor outputs and inputs that are entered by the operators. Hence, we developed a new inference engine in ExpressIF[®], called dialog engine. Indeed, the engine is able to fire the rules that can be fired, i.e whose inputs are based on sensors. When an input is missing, its value is asked to the operators. This is why we talk about dialog, because the engine is able to stop the inference, ask some values to the user, resume the inference, and loop until a decision is made.

Not all the missing values are necessary to make a decision. For instance, in the case of a conjunctive premise, if one of the proposition is evaluated to 0, all the premise will be null. In that case, the engine will not ask the user the missing values because they are not necessary for the decision making process. This gives the illusion of a smart dialog.

The dialog engine is extensible but for now, it supports different kinds of discussion fragments:

- the user can send inputs values, ask for explanations;
- the engine can send the output values or explanations, ask for missing inputs.

3.7 Summary

In this first chapter, we explored the first family of contributions that aims at improving fuzzy systems as XAI by increasing their expressiveness. The extensibility of our system ExpressIF[®] allows adding seamlessly new predicates and new operators and grouping them by domains. We presented the main three domains to which we contributed: temporal, spatial and spatio-temporal. For instance, we have also predicates to compare phonetically or syntactically strings.

These predicates are represented by fragments of sentences that can be used in the knowledge base. This improves its interpretability since it reduces the distance between the formalized knowledge and the expert knowledge. In addition, it also increases the number of problems that can be tackled by the system. Finally, they can also be used during the generation of textual explanations.

Creating a new predicate is thus the matching between a part of sentence and an algorithm to assess its value. The most difficult part is that the predicates have to be evaluated from a variety of data: signals, images, DTM, scalars etc. Technically, we have succeeded in allowing ExpressIF[®] reasoning on all those data and dimensions. We have also presented the compositional paradigm that is used to define predicates from base predicates.

Finally, we also describe the links between predicates as a piece of knowledge available for other algorithms. We will show in the next chapter how this can be used, in particular in the case of predicate learning (see chapter 4, section 4.3.3).

Chapter 4

Extracting knowledge from data for building fuzzy systems

4.1	Motivations	72
4.2	Interpretable feature construction	74
4.2.1	Context	74
4.2.2	Feature construction	75
4.2.3	Interpretable feature construction as a prior method	76
4.2.4	Interpretable embedded feature construction	79
4.2.5	Evaluation of the interpretability of constructed features	80
4.2.6	Results of DVCS event selection	84
4.3	Relevant predicate learning	85
4.3.1	Generalization versus specialization	85
4.3.2	Extraction of relevant predicates	86
4.3.3	Heuristics to decrease the number of predicates to evaluate	88
4.3.4	Predicate learning for image classification	88
4.3.5	Predicate learning for image object or region annotation	89
4.3.6	Application to medical image annotation	90
4.4	Extraction of gradual rules from experimental data	92
4.4.1	Context	92
4.4.2	Causality and graduality extraction	93
4.4.3	Representation and evaluation of gradual rules	94
4.4.4	Validation on a toy dataset	96
4.4.5	Application to materials properties prediction	97
4.5	Summary	99

Until 2016, ExpressIF[®] was used only to model expert knowledge: it was called the "no-data approach". We were not ignoring data proliferation or the importance of machine learning (ML). We were focusing on other topics and applications. Nevertheless, from 2016, we started looking at the state-of-the-art methods to compare them and try to apply them on our real-world data. In this chapter, we present three contributions that allow building fuzzy systems that are interpretable and explainable.

The first contribution was born from an observation: interpretable models (e.g. decision trees, fuzzy or not, Generalized Additive Models, etc.) do not build their own representation of data, contrary to neural networks, in particular deep neural networks. It may

explain their lower performances against black box models (Sondhi, 2009), which are obviously less transparent. In addition, we worked before on new manually built features for a certain type of sensors, which emphasizes the importance of good features (see section 6.2). We proposed to build new features from the original ones. Feature construction is not new, but the novelty is to enforce the interpretability and the mathematical plausibility of such features. This is one of the results from Noëlie Cherrier’s PhD (Cherrier, 2021).

The second contribution is more about seizing an opportunity. Since ExpressIF[®] can manipulate a large vocabulary (see chapter 3), why not take advantage of it to build more interpretable and explainable models. We thus focus on spatial relation learning for classification or semantic annotation. This was carried out during Regis Pierrard’s PhD (Pierrard, 2020).

The last contribution is a proof of concept in the context of materials science. We address the problem of the extraction of relevant knowledge, specifically gradual rules, from experimental data. These rules are valuable knowledge for the researchers but it is also a tool to approximate monotonous functions. This was investigated during Hiba Hajri’s postdoctoral fellowship and Killian Susini’s internship.

Before describing some contributions about symbolic learning, I will start motivating work.

4.1 Motivations

Among all the literature algorithms that we have implemented in ExpressIF[®], fuzzy decision trees and FURIA (Fuzzy Unordered Rule Induction Algorithm) (Hühn and Hüllermeier, 2009a) are the most popular methods. In our experience, by far, FURIA is the fastest rule induction algorithm and the one that can be applied on various use cases with the best performances. However, the principles of FURIA are quite simple since it is a fuzzification of the crisp rule learning algorithm RIPPER (Repeated Incremental Pruning to Produce Error Reduction) (Cohen, 1995). It aims at producing unordered rule sets instead of decision lists and at fuzzifying the crisp boundaries learned. FURIA has also an optimization phase to avoid overfitting. After training, FURIA rule base is used for classification, and three cases arise:

- Only one rule is activated more than the others. It is then used to affect the label.
- Two or more contradictory rules are activated. In this case, FURIA weights the rules regarding their certainty factor.
- The instance is not covered by any rule. FURIA executes a process called "rule stretching". A stretching of a rule consists in removing antecedents, and it is said minimal if it does not remove more antecedents than necessary to cover the current instance.

Let us focus on the third case. How many instances are not covered by FURIA rules? To evaluate that, we proceeded to a 5-fold cross validation using FURIA on several famous datasets, deactivating the rule stretching procedure. Table 4.1 shows the results in terms of correct and incorrect classification of FURIA applied to different datasets. In addition, it shows the number of instances that cannot be classified without rule stretching. With rule stretching, all the instances can be classified and generally, the number of instances that are correctly classified increases.

My interpretation is that the rules extracted by FURIA are not "valuable": they do not reflect the phenomenon behind the data. Actually, their expressiveness is limited because RIPPER uses thresholds that FURIA fuzzifies.

Dataset	Rules	Correct classification (%)	Incorrect classification (%)	Unclassified instances (%)
Breast cancer	3.8 ± 0.8	69.2 ± 4.4	23.8 ± 5.2	7.0 ± 6.0
Diabetes	6.8 ± 5.2	70.8 ± 6.1	21.6 ± 1.9	7.5 ± 7.0
Glass	11.4 ± 1.8	64.4 ± 4.0	20.1 ± 5.3	15.9 ± 6.0
Credit	10.4 ± 4.8	43.6 ± 2.5	20.1 ± 1.2	36.3 ± 2.6

Table 4.1: Results of FURIA, obtained with a 5-fold cross-validation on several famous datasets. The classification results are given as the mean \pm standard deviation over the folds.

The problem is almost the same with fuzzy decision trees. In their simplest form, their expressiveness resides in the linguistic variables. Nevertheless, they generally do not have unclassified instances because strong partitions are used.

The motivations of this part of my work come from this. The goal is to extract valuable knowledge from data. It is not referring to linguistic summaries of datasets (Shukla et al., 2020; Smits et al., 2018), even if it borrows some algorithms from it. It rather refers to a kind of knowledge that can be executed to make decisions (with high performances) and that also brings insights to a human user, to help understanding the phenomenon that is studied. If we want users to understand these insights, the vocabulary must be appropriate and rich enough to express them. Hence, this is also related to interpretability and explainability.

I give my own definition of interpretability and explainability, related to the field of XAI.

Definition 4.1 (*Interpretability*)

The interpretability is a capacity of a model to be interpretable, i.e. understood by a user, in the sense that this user is able to make statements about its behavior. This understanding may be direct or by means of tools: the degree of interpretability is thus related to the complexity of the tools that help understanding the model.

The definition of interpretability is thus related to a user and his/her abilities. The same model can thus be interpretable for a given user and not interpretable for another user.

Definition 4.2 (*Explainability*)

The explainability is a capacity of a model to provide a user with an explanation of the decision. An explanation is a set of clues whose natures depend on the task accomplished by the model, the application domain and the user for whom it is intended.

There is thus a difference between interpretability and explainability since the first concerns the comprehension of the model and the second the comprehension of the decision.

As stated by Alonso (Alonso Moral, Castiello, et al., 2021), if a model is interpretable, it may be easy to make it explainable. Otherwise, if the model is not interpretable, the explainability may be difficult to obtain, or even impossible.

The next sections are thus dedicated to contributions about interpretability and symbolic machine learning.

4.2 Interpretable feature construction

4.2.1 Context

Since antiquity, scientists investigated the structure of matter in greater and greater detail, as illustrated in figure 4.1. The goal of High Energy Physics (HEP) is to determine the most fundamental building blocks of matter, the subatomic or elementary particles, and to understand the interactions between them. The general context of Noëlie Cherrier’s PhD was the study of the proton structure.

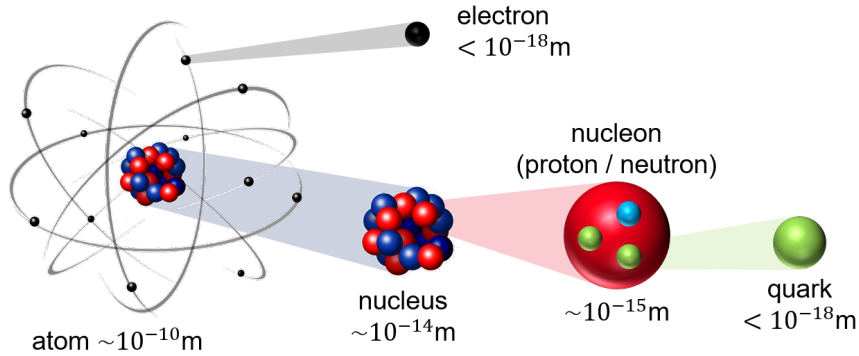


Figure 4.1: Inside of an atom (Cherrier, 2021).

Broadly, elementary particles and the way they interact with each other are studied in particle collider facilities. In our case, it is the CLAS12 experiment, at the Jefferson Laboratory (Newport News, Virginia, USA). Out of the collision between two particles, new particles emerge and interact with the detectors located around the collision site. We are interested in the discrimination of two types of interaction:

- The Deeply Virtual Compton Scattering (DVCS) denotes the scattering of an electron off a proton accompanied by a photon: $ep \rightarrow ep\gamma$.
- The π^0 electroproduction results in a π^0 instead of the photon that decays into two photons: $ep \rightarrow ep\pi^0 \rightarrow ep\gamma\gamma$.

The CLAS12 experiment has been designed to record DVCS events in a large phase space. The DVCS cross-section is directly proportional to the number of observed DVCS events in CLAS12, the other factors being the luminosity and CLAS12 acceptance among others. The experimental goal is therefore to isolate DVCS events from all CLAS12 data. Two cases may therefore compromise the DVCS selection:

- a DVCS event that has an additional photon for various reasons (radiation of the electron for instance) may be confounded with a π^0 electroproduction event;
- a π^0 electroproduction event in which one of the two photons has a very high energy and the other one very low may be taken for a DVCS event, because the lowest energetic photon has probably remained undetected.

Therefore, a physics analysis necessarily involves a careful event selection to retrieve the maximum number of DVCS events while minimizing the π^0 contamination. We compared different approaches: a traditional mathematical analysis, a neural network (the topic of Marouen Baalouch’s postdoctoral fellowship) and transparent models with automatic feature construction. In this section, I focus on the latter but the whole comparison is available in (Cherrier, 2021). The use of transparent models in sciences is motivated by the fact that the validation of an analysis method is a *sine qua none* condition for publication authorization for acceptance at the peer-reviewing stage.

The CLAS12 dataset gathers up to five particles of the output state to form the feature set: one electron, one proton, and up to three photons ranked by missing mass $ep \rightarrow ep\gamma$. In total, the 35 available features are the three-dimensional momentum (namely mass, times, speed of the particle) for each identified particle, expressed in two coordinate systems, and the three-vector as-is: $p_x, p_y, p_z, p_T, \theta, \phi, \mathbf{p}$.

4.2.2 Feature construction

Feature engineering refers to any processing of the feature space in order to constitute an efficient feature set as input to machine learning algorithms. It gathers a large range of techniques, including feature extraction from raw data, data encoding, and feature transformation among others.

Feature selection and feature construction are two related subfields of feature engineering. On the one hand, the objective of feature selection is to reduce the dimensionality of the feature set, notably to avoid overfitting. On the other hand, feature construction aims at building new features from the original ones, so that the performance of a machine learning algorithm is improved. In other words, feature selection will never make up for an initial inappropriate feature set, while feature construction directly searches for new relevant features for the learning task.

Regarding the classification algorithm and the method to evaluate a candidate feature, both feature construction and feature selection methods can be divided into three categories (Kohavi, John, et al., 1997):

- *Filter methods* compute a candidate feature score independently of the ML algorithm. However, the final feature set may not be the most adapted regarding the chosen ML algorithm since the latter was not involved in the process.
- *Wrapper methods* use the prediction score of a ML algorithm to evaluate the candidate features. The main drawback of this class of methods is the computation time. However, the final feature set may be more specific to the ML algorithm.
- *Embedded methods* combine the feature construction/selection process with the training of the ML model. These methods are usually fast but very specific to the chosen ML algorithm and hardly generalizable.

More details can be found in (Cherrier, 2021). We have chosen to use a genetic programming (GP) approach, i.e. an evolutionary computation technique comparable to genetic algorithms. Individuals in GP are represented by trees instead of vectors as in genetic algorithms (Koza, 1992). Figure 4.2 illustrates the evolution process: a generation n of individuals undergoes crossover (exchanging branches) and mutation (modifying a branch). The new individuals constitute what is called the "offspring" and are evaluated. Finally, a selection is performed to obtain the next generation. In our opinion, GP is the most flexible feature construction method for incorporating domain knowledge and notably for enforcing feature interpretability.

This research is motivated by the observation that datasets involve quantities that are associated with a dimension and a physical unit. When combining these base features into high-level variables through automated feature construction, some associations must be forbidden: for instance, adding a length and a weight makes no sense. Moreover, each application field may have some recurrent patterns that can be exploited to guide the search for relevant high-level features. The underlying idea is to reproduce the approach of creating high-level features, common in fundamental sciences. For instance in High-Energies Physics, variables such as the missing masses or invariant masses are frequently used to select signal events, and consist in the square root of a sum of squared quantities (actually the norm of a four-momentum).

Therefore, two forms of prior knowledge can be used:

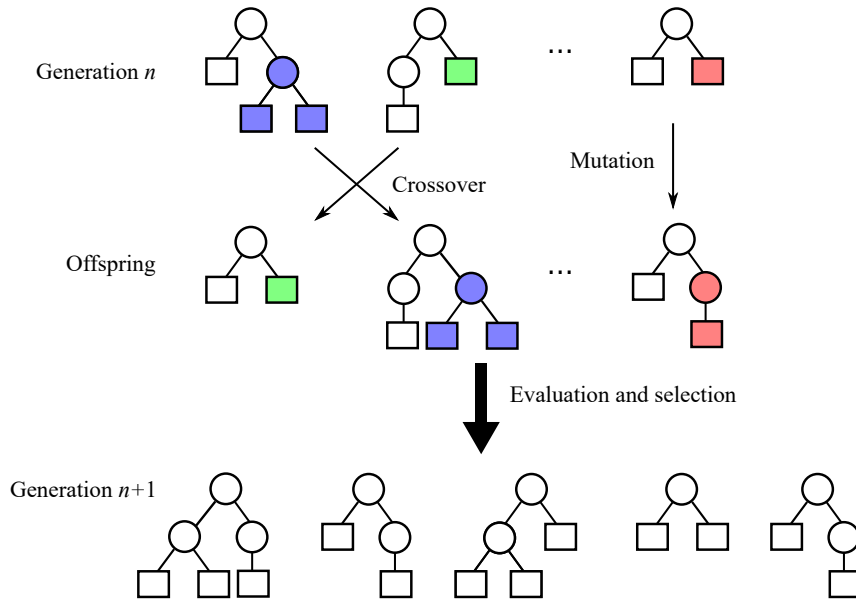


Figure 4.2: Overview of the principle of genetic programming (Cherrier, 2021).

- the knowledge of the base variables units and dimensions;
- the experience of their usual combinations in the application domain.

These two forms of guidance allow obtaining interpretable features since unit and dimension-consistent features are obviously mathematically valid and since features that resemble in some ways to usual variables used by domain experts are easier to read.

In the next subsection, we describe our approach to interpretable feature construction as a prior method.

4.2.3 Interpretable feature construction as a prior method

In this subsection, we describe our method to build new features before the training of any ML algorithm. It results in adding new columns to an existing dataset.

Grammar-based unit and dimensionality constraint

Grammar-based GP permits to enforce constraints during feature construction. Following the idea of (Ratle and Sebag, 2001), we defined a context-free grammar that constrains feature construction so that only valid features can be built from the point of view of units and dimensions. We limited the achievable unit powers so that the algorithm does not produce quantities with unit cm^8/kg^3 for instance. This grammar differs a bit depending on the dataset. An example of such grammar for the CLAS12 dataset is given in figure 4.3.

Transition matrix for guiding towards usual combinations

The default grammar-based GP algorithm randomly selects a production rule each time a non-terminal symbol must be transformed. Usually, regarding an application domain, some operators are more frequent than others. For instance, in physics, division is not often employed contrary to addition. This is logical because physics analyses largely rely on energy and momentum conservation, i.e. summing the input and subtracting the output.

To enforce our own guidance on the feature search and favor the construction of formulas that are similar to those used in the application domain, we chose a probability distribution on the transitions between operators. For instance, a square root in physics is very often

$\langle \text{start} \rangle$::= $\langle \text{E} \rangle$ $\langle \text{E2} \rangle$ $\langle \text{A} \rangle$ $\langle \text{F} \rangle$
$\langle \text{E} \rangle$::= $\langle \text{E} \rangle + \langle \text{E} \rangle$ $\langle \text{E} \rangle - \langle \text{E} \rangle$ $\langle \text{E} \rangle \times \langle \text{F} \rangle$ $\langle \text{E} \rangle \div \langle \text{F} \rangle$ $\text{sqrt}(\langle \text{E2} \rangle)$ $\text{norm}(\langle \text{M} \rangle)$ $\langle \text{component} \rangle(\langle \text{M} \rangle)$ $\langle \text{termE} \rangle$
$\langle \text{E2} \rangle$::= $\langle \text{E2} \rangle + \langle \text{E2} \rangle$ $\langle \text{E2} \rangle - \langle \text{E2} \rangle$ $\langle \text{E2} \rangle \times \langle \text{F} \rangle$ $\langle \text{E2} \rangle \div \langle \text{F} \rangle$ $\langle \text{E} \rangle \times \langle \text{E} \rangle$ $\text{square}(\langle \text{E} \rangle)$ $\text{dot}(\langle \text{M} \rangle, \langle \text{M} \rangle)$ $\langle \text{termE2} \rangle$
$\langle \text{A} \rangle$::= $\langle \text{A} \rangle + \langle \text{A} \rangle$ $\langle \text{A} \rangle - \langle \text{A} \rangle$ $\langle \text{A} \rangle \times \langle \text{F} \rangle$ $\langle \text{A} \rangle \div \langle \text{F} \rangle$ $\text{angle}(\langle \text{M} \rangle, \langle \text{M} \rangle)$ $\langle \text{termA} \rangle$
$\langle \text{F} \rangle$::= $\langle \text{F} \rangle + \langle \text{F} \rangle$ $\langle \text{F} \rangle - \langle \text{F} \rangle$ $\langle \text{F} \rangle \times \langle \text{F} \rangle$ $\langle \text{F} \rangle \div \langle \text{F} \rangle$ $\langle \text{E} \rangle \div \langle \text{E} \rangle$ $\langle \text{E2} \rangle \div \langle \text{E2} \rangle$ $\langle \text{A} \rangle \div \langle \text{A} \rangle$ $\text{sqrt}(\langle \text{F} \rangle)$ $\text{square}(\langle \text{F} \rangle)$ $\text{cos}(\langle \text{A} \rangle)$ $\text{sin}(\langle \text{A} \rangle)$ $\text{tan}(\langle \text{A} \rangle)$ $\langle \text{termF} \rangle$
$\langle \text{M} \rangle$::= $\langle \text{M} \rangle + \langle \text{M} \rangle$ $\langle \text{M} \rangle - \langle \text{M} \rangle$ $\langle \text{termM} \rangle$
$\langle \text{component} \rangle$::= get_x get_y get_z

Figure 4.3: Grammar used for the CLAS12 dataset. E stands for a 1D momentum or energy in GeV, E2 for a squared momentum or energy in GeV^2 , A for an angle in radians, F for a unitless real number, M for a three-momentum of unit GeV. $\langle \text{termX} \rangle$ means a terminal of type X, namely a base feature or a constant.

followed by a sum of squares. In this way, we also forbade a square operator followed by a square root and conversely, to simplify the trees. This last constraint could actually appear in the grammar itself, but it would lead to more complex and less readable grammar. It has no importance if the probabilities used are not accurate, but they must reflect the true usual combinations.

The probabilities may thus be given by an expert. An attempt has been made to extract these probabilities automatically from documents: articles, books, lectures, etc. This was developed with Sébastien Klasa during his second year of apprenticeship. However, notations were often simplified and were mostly using vector notations instead of separate components. Therefore, this attempt to automate the design of the transition matrix highlighted the need for specific research in natural language processing adapted to scientific documents. An example of probabilities for the CLAS12 dataset is given in table 4.2.

Workflow of GP based feature construction

From the grammar and transition matrix, one can obtain valid tree-like individuals forming a population that can be evolved with GP.

We considered both single and multiple feature construction: in the multiple feature construction case, a multi-tree representation is used and therefore evolves a population of individuals, which are lists of trees.

An initial population is first generated, evaluated and then evolved. For each individual in the population, mutation can be applied with a probability P_{mutation} and crossover with a probability $P_{\text{crossover}}$. The offspring is then evaluated, and the selection is performed over the whole offspring and parent population, which has the advantage of keeping in the population some efficient features from the parent population.

To build a new tree, a type T and the first operator (returning type T) are selected under the initial probability distribution. Then, while the condition on the current depth is not satisfied, the tree keeps growing. The possible operators are selected according to the grammar. The transition matrix then defines a distribution probability among the possible operators according to the parent node. Finally, the leaves of the tree are randomly chosen among the set of base features and constants of the proper type.

After the population is initialized, the evolution process starts with a series of mutations, crossovers, evaluations and selections. However, the generation technique is still used

Return type: {E: 0.5, A: 0.2, F: 0.1}.							
	E + E	E - E	E × F	E ÷ F	sqrt(E2)		
E + E	0.1	0.1	0.1	0.1	0.6		
E - E	0.225	0.225	0.25	0.2	0.1		
	E + E	E - E	E × F	E ÷ F	norm(M)		
square(E)	0.5	0.1	0.07	0.03	0.3		
	E2 + E2	E2 - E2	square(E)	E × E			
E2 + E2	0.4	0.15	0.4	0.05			
E2 - E2	0.2	0.07	0.7	0.03			
sqrt(E2)	0.7	0.25	0	0.05			
	E + E	E - E	E × F	E ÷ F	sqrt(E2)		
E × F	0.15	0.15	0.35	0.3	0.05		
E ÷ F	0.15	0.15	0.35	0.3	0.05		
	F + F	F - F	F × F	square(F)	cos(A)	sin(A)	tan(A)
E × F	0.025	0.025	0.025	0.025	0.375	0.375	0.15
E ÷ F	0.025	0.025	0.025	0.025	0.1	0.1	0.7
		M + M	M - M				
	norm(M)	0.9	0.1				

Table 4.2: Transition matrix for the CLAS12 dataset. The probabilities are displayed for the next possible operations (as columns) given the previous one (as row). Operations that are not listed as rows have a uniform transition probability distribution. Operations that are not listed as columns for a given previous operation cannot be selected as the next operation (probability 0). The notations are the same than in the grammar (Figure 4.3).

during the evolution each time a tree or a subtree needs to be created, to keep following the same grammar rules and transition probabilities.

Mutation and crossover operators apply classically at each generation of the GP algorithm to constitute the offspring.

The mutation method is randomly picked among three existing techniques. Each of these techniques is modified to be compatible with the grammar:

- Uniform mutation: a node is selected in the tree and then the subtree is entirely regenerated from that node while making sure that the dimensional consistency is still respected in particular at the root.
- Node replacement: a node is selected and replaced with any dimensionally compliant node.
- Insertion: from a selected node, a new subtree is inserted that has the original subtree(s) of the mutated node as child nodes. The inserted subtree is generated so that the grammar is also respected at the connections with the original tree.

The transition matrix is used each time a new tree or subtree needs to be generated to support the interpretability. In the case of multiple feature construction, mutation is applied on each tree of the list.

The crossover operation is the standard GP one-point crossover, assuming that the exchange of the two subtrees is compatible with the grammar, i.e. that the two roots of the subtrees share the same type. For multiple feature construction, the crossover is applied on the lists instead of on the trees: two sublists are exchanged.

To evaluate an individual, trees are converted to numerical features by computing the function they represent on the base features. Any invalid operation (e.g. division by zero)

creates a missing value. Then, a fitness function must evaluate the transformed individual. To this end, several methods exist. For instance, wrapper methods evaluate the score (e.g. the accuracy) of a classifier trained with the newly built feature(s), while filter methods use ranking measures that are independent of any predictor.

The selection applies to the joint set of individuals from the offspring and from the parent population. A repeated tournament selection among three randomly picked individuals constitutes the next generation population. The tournament size of three is a compromise between randomness and quick convergence.

To emphasize the interest in building a new feature, we made a simple experiment with the construction of one feature. Table 4.3 compares the kappa scores of several classifiers on the original dataset and the same dataset extended with one built feature. All results are presented with their mean and standard deviation over at least 25 independent runs (5 for each fold, for 5 folds). Whatever the classifier, the score is always improved.

	Baseline	With 1 built feature
Fuzzy C4.5	0.345 ± 0.019	0.396 ± 0.020
CART	0.243 ± 0.016	0.294 ± 0.025
AdaBoost	0.333 ± 0.019	0.361 ± 0.017
GradientBoosting	0.302 ± 0.012	0.396 ± 0.020
FURIA	0.236 ± 0.011	0.311 ± 0.030
GAM	0.320 ± 0.006	0.391 ± 0.014

Table 4.3: Cohen’s kappa score for one built feature with our method and information gain as fitness function, using different classifiers for evaluation, performed on CLAS12 dataset.

Among the most common built features, some have a straightforward interpretation:

- $p_z^e + p_z^p + p_z^{\gamma_1}$ is related to the momentum conservation for the z component. In the case of a DVCS, this value should be approximately 10.6 GeV/c. Thus, it is no wonder it has a great discriminative power.
- $\text{angle}(\vec{p}^{\gamma_1}, \vec{p}^{\gamma_1} + \vec{p}^{\gamma_2})$ is equivalent to the angle $\text{angle}(\vec{p}^{\gamma_1}, \vec{p}^{\pi^0})$ (regarding the momentum conservation again). That can be interpreted by the fact that this angle should be as small as possible, because the decayed photon should be close to its mother particle.

This confirms that our automatically built features have an interpretation in physics.

4.2.4 Interpretable embedded feature construction

Instead of using the constrained feature construction method presented before as a prior step before training a classifier, this subsection focuses on embedding the feature construction in the induction process of the classifier. This approach uses a filter fitness function, quicker to evaluate than a wrapper one. The constrained feature construction algorithm has been implemented in tree-based (Cherrier, 2021; Cherrier, Poli, Defurne, and Sabatié, 2020) and rule-based algorithms and in Generalized Additive Models (Cherrier, 2021; Cherrier, Mayo, et al., 2020) with the goal to include prior knowledge. In this document, we limit ourselves to the embedding of feature construction in tree-based and sequential covering algorithms only.

Principle for tree-based models induction

For tree-based models, one feature can potentially be built at each node of the tree(s). Just as ensemble methods restrict the complexity of individual classifiers, the number of

feature constructions is limited and the construction algorithm is restrained. A parameter N_{max} controls the maximum number of features that it is authorized to build. When the number of allowed constructions is restrained, the features are built from the root and level by level, going down as the tree is formed. Such principles may be adapted to different tree-based algorithms: C4.5, CART, fuzzy decision trees, tree ensemble methods such as adaptive boosting or gradient boosting.

Generally, they are based on a discrimination measure (e.g. information gain in C4.5) that will be used as fitness function during the feature construction. In the particular case of tree ensembles, the number of built features is already large if $N_{max} = 1$ for each individual tree. Thus, the parameter N_{max} is altered in this case to admit a probability to build a single feature at the root of the tree. It does not control in which tree a feature is built or not.

Principle for rule base induction by sequential covering

In sequential covering algorithms, a feature construction can be performed at each addition of an antecedent in a rule. As a reminder, sequential covering algorithms progressively add rules so that the training set is covered with a minimal set of rules. In RIPPER and FURIA for instance, antecedents are added to a rule to maximize the FOIL (First-Order Inductive Learner) information gain (Hühn and Hüllermeier, 2009b). In the same way as for decision trees, a feature construction algorithm can be performed at this stage, the fitness function being the discrimination measure i.e. the FOIL’s information gain.

A rule base is treated the same way as a tree ensemble: N_{max} is the number of built features per rule. To allow for a smaller number of built features in total, N_{max} can take values below 1 to reflect the probability to build a single feature in a given rule. There is no consideration of the order in which the rules are built.

With the same dataset as the section before, we performed some tests to compare the performances of the embedded feature construction in C4.5 and FURIA algorithms with the base algorithms (called baselines). The results are shown in table 4.4.

	Baseline	With FC (number of built features)
Fuzzy C4.5	0.345 ± 0.019	0.476 ± 0.025 (15)
FURIA	0.236 ± 0.011	0.496 ± 0.039 (2)

Table 4.4: Cohen’s kappa score for built features with our embedded method and information gain as fitness function, using different classifiers for evaluation, performed on CLAS12 dataset. FC stands for feature construction. The numbers between parentheses are the number of built features.

4.2.5 Evaluation of the interpretability of constructed features

Protocol

There are different ways to evaluate the interpretability of features built automatically by the feature construction algorithm. For instance, in the case of the features built for the CLAS12 dataset, we can ask physicists to analyze them.

In addition, since interpretability must be evaluated by the target users of the proposed models, we conducted an application-specific experiment with experimental physicists. According to the classification of Doshi *et al.* (Doshi-Velez and Kim, 2017), our study was *application-grounded*: it concerned humans (here physicists) and real tasks (here the classification of events). However, (Doshi-Velez and Kim, 2017) state that large-scale

experiments are difficult to conduct since target users are experts of the field, therefore not numerous. Besides, human-produced explanations make a good baseline.

In this study, we designed a survey targeted to physicists to evaluate their perceived interpretability of the different features, automatically built by feature construction.

Considering the randomness of the feature construction process on the one hand, and the recommendation of (Doshi-Velez and Kim, 2017) to compare setups by pairs on the other hand, we established the following evaluation protocol. First, we determined a list of pairwise matches that we wanted to make:

- our feature construction method against the standard unconstrained GP-based technique;
- feature construction prior to model induction against embedded feature construction;
- automatic feature construction techniques (prior or embedded) against regular variables used by expert physicists.

For each of these three matches, we randomly picked four features for each of the two involved categories, leading to eight features. The respondent was then asked to give a mark between 1 and 5 to each feature, from poorly understandable (1) to highly understandable (5), according to his perception of the physical meaning and of the relevance of the proposed feature. The respondent was encouraged to use the full range of available marks, since the marks were independent between each successive match. In total, the respondents rated 24 features (8 for each of the three matches).

The marks obtained by each group were compared globally among all responses. If the difference was significant, then we concluded that one group was significantly more interpretable than the other.

Evaluation

We asked physicists to participate in the pool we described before. We obtained from 24 to 31 answers. The distributions of age and professional status properly covered all categories of the researchers population. 87.5% of the respondents declared being at least curious about machine learning and 50% believed that using machine learning for the full reconstruction-analysis chain would significantly improve the physics output.

Let us first compare constrained against unconstrained feature construction. Figure 4.4 displays the results of the block of features containing four features built with a constrained GP algorithm and four features built with the unconstrained GP algorithm. It is clearly visible that constrained features have been better rated than the unconstrained ones: the average score for the unconstrained feature construction is 1.69, against 3.19 for constrained feature construction. One exception subsists though: the last feature has been built with constrained GP but is badly noted. The feature is:

$$\left(p_T^p + \left\| \overrightarrow{p^{\gamma_2}} \right\| \right) \tan \left(\text{angle} \left(\overrightarrow{p^{\gamma_2}}, \overrightarrow{p^{\gamma_1}} \right) \right). \quad (4.1)$$

Although it respects the physical units, we can guess that the sum of a transverse momentum with a norm is complex to apprehend and to understand.

Figure 4.5 displays the results of the block of features opposing prior and embedded feature construction. No clear tendency emerges from the visualization of the results: the average score for embedded feature construction is 2.86, closely ahead of prior feature construction with 2.77. Performing feature construction in an embedded way does not seem to impair interpretability, while it is computationally more efficient than prior feature construction. However, our intuition is that four features are not enough to be representative of embedded feature construction: this category covers many algorithms (crisp and fuzzy

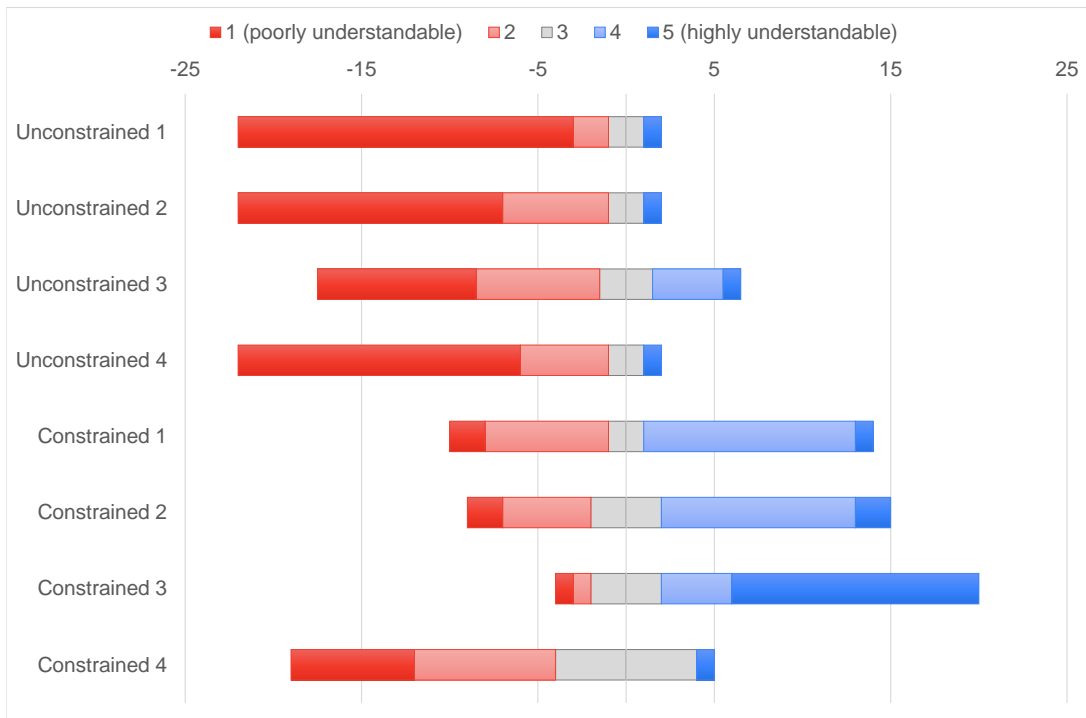


Figure 4.4: Results of the survey on constrained against unconstrained feature construction (FC). The distributions of the responses are displayed for each feature.

decision trees, FURIA, GAM) and features built at different levels of specificities. Indeed, a feature built at the root of a decision tree has a high probability to be similar to a feature built with prior feature construction since it uses all available data. However, a feature built in a deeper node of the tree is specific to the data subset that reaches this node. Intuitively, the more specific the features, the less understandable.

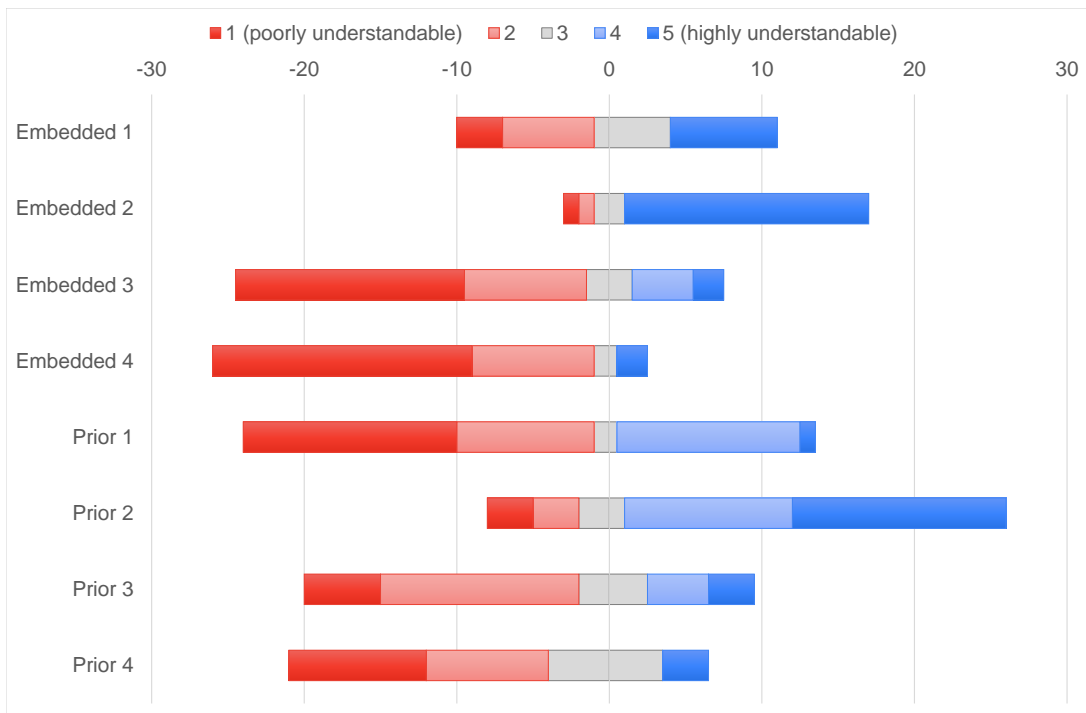


Figure 4.5: Results of the survey on prior against embedded feature construction (FC). The distributions of the responses are displayed for each feature.

Figure 4.6 presents the results of the last comparison, namely automatic feature construction against regular variables used by physicists. Again, there is no significant difference between the two categories, but physicists' regular variables are still slightly ahead of automatically built features, with an average score of respectively 3.74 and 3.51. However, the score obtained by the two last physicists' regular variables (features 7 and 8 on Figure 4.6) must be discussed. Indeed, they are commonly used by physicists for exclusivity cuts, but were written in the survey as mathematical formulas that might have not been recognized:

- missing mass $ep \rightarrow e\gamma X$:

$$\sqrt{\left(-\|\vec{p}^{\tilde{e}}\| - \|\vec{p}^{\gamma 1}\| + M_p + e_z^{in}\right)^2 - \left\|-\vec{p}^{\tilde{e}} - \vec{p}^{\gamma 1} + \vec{e}^{in}\right\|^2}, \quad (4.2)$$

- missing energy $ep \rightarrow ep\gamma X$:

$$e_z^{in} + M_p - \|\vec{p}^{\tilde{e}}\| - \sqrt{\|\vec{p}^{\tilde{p}}\|^2 + M_p^2} - \|\vec{p}^{\gamma 1}\|. \quad (4.3)$$

This last feature has indeed received contradictory scores, with 23% of respondents saying it is poorly understandable and 32% that it is highly understandable. With textual formulations such as "missing mass" and "missing energy", probably these features would have received a higher score. However, comparing all features on equal terms indicates that features with complex mathematical formulations are considered less understandable.

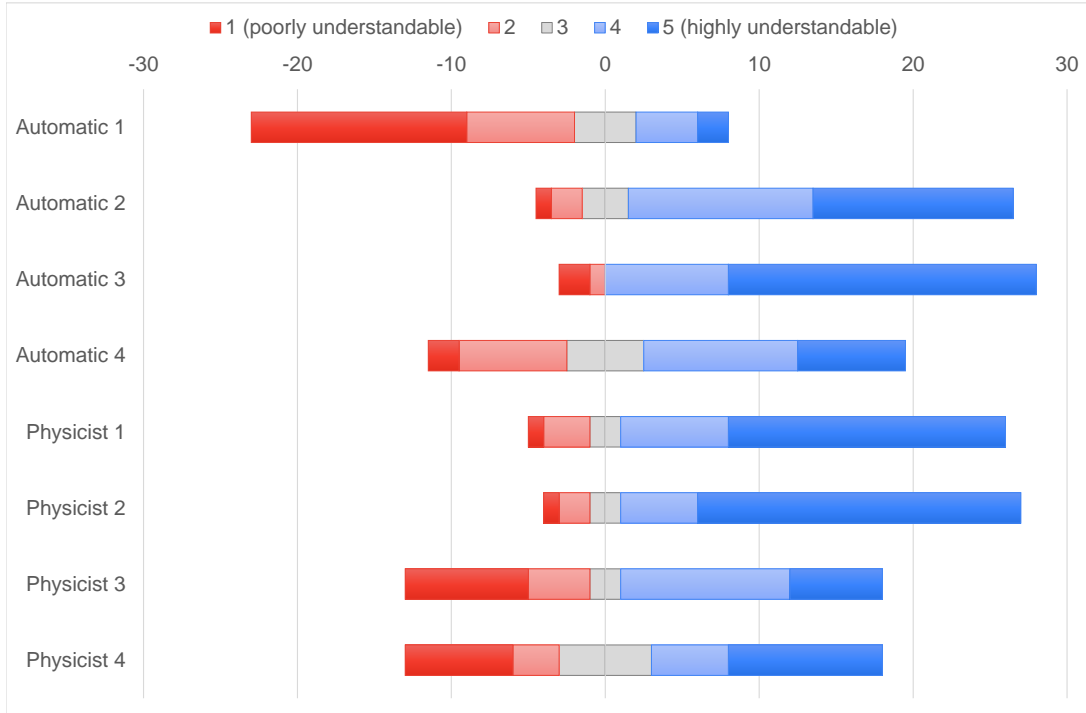


Figure 4.6: Results of the survey on automatic feature construction (FC) against regular variables used by physicists. The distributions of the responses are displayed for each feature.

Globally, there is no significant difference between the understandability of automatically built features and physicists' regular variables. Therefore, our proposed constrained feature construction algorithm permits the automation of the feature engineering step without significant loss in interpretability. Moreover, the automatically built features were demonstrated more discriminant than physicists' regular features.

Finally, we drew a few conclusions from this survey:

- The constrained feature construction algorithm (PGGGP) produces significantly more understandable features than the unconstrained (GP) version. This was quite obvious objectively since the majority of the built features with the GP algorithm do not respect the physical units, but this is now proven subjectively.
- No clearly visible difference appears in the understandability of features built prior to model induction or embedded into it, but this could be due to a lack of representativity notably for embedded feature construction.
- Regular physicists’ variables are probably noted slightly more understandable than automatically built features, but it must be emphasized that the mathematical formulas of automatically built features are often more concise, which is an asset towards understandability.

4.2.6 Results of DVCS event selection

I do not go into the details of all the processing steps that have been applied on the CLAS12 dataset. They are related to the physics part of the project. We performed a comparison between 5 approaches:

- a rule base with 11 rules from FURIA, with prior construction of 5 features;
- a fuzzified version of C4.5 with embedded feature construction of 15 features;
- a Generalized Additive Model with 16 terms and feature construction (FCGAM), and bitonic constraint as proposed in (Cherrier, Mayo, et al., 2020);
- a neural network from the post-doctoral fellowship of Marouen Baalouch (Baalouch et al., 2019);
- physicists’ cuts from the PhD of Guillaume Christiaens (Christiaens, 2021), representing the typical approach in HEP.

To compare the models, we looked at the statistics (i.e. number of selected events) and the contamination percentages. It is usual in HEP to divide the space into non-homogeneous bins, called kinetic bins. In our case, the usual physicist analysis was performed on 24 bins. Table 4.5 shows the results in three representative bins: bin 3 comprises many contaminating events, bin 17 almost none, and bin 2 is an intermediate.

	Bin 2		Bin 3		Bin 17	
	Stat.	Cont.	Stat.	Cont.	Stat.	Cont.
Cuts	5434	10.4%	11517	33.6%	12459	3.8%
Neural network	6287	8.5%	15450	30.9%	14456	3.5%
FURIA	8148	12.1%	20211	37.6%	15649	5.6%
FCGAM	6648	9.1%	15743	30.3%	14252	3.4%
Fuzzy C4.5	5588	9.1%	15293	32.0%	13972	3.4%

Table 4.5: Statistics and contamination percentage in three bins with several models.

FURIA was systematically obtaining the larger number of selected events, at the expense of a higher contamination level. The neural network was having the lower contamination level in bin 2, while it was the FCGAM in bin 3. Two models were tied in bin 17: the FCGAM and the fuzzy C4.5. The FCGAM was the one retaining the most statistics. Overall, excluding FURIA, the cuts had the lowest statistics and the highest contamination in all bins. This proved the great interest of machine learning for a physics analysis:

for instance, using the FCGAM permitted to increase the statistics by 24% in average over these three bins, while reducing the contamination.

FURIA was the only model surpassing all the others. However, these results should be taken with very high precaution: indeed, FURIA was keeping much more events than the other models, including 13% of identified η production events. Therefore, its performance must be read with caution since there was a high associated systematic error, linked to other background processes that were not properly removed. Thus, FURIA was not applicable for this particular analysis.

We also surveyed the interpretability of the models, with the same panel as before asking how a given example was classified, how to modify the model to change the classification, how sure was the respondent about his answers and a subjective evaluation of the model's transparency or ease of validation. The proposed FCGAM turned out to be a very satisfying model: it was rated as the preferred model in terms of interpretability by the physicists and obtained similar or better performances than the neural network for the data analysis.

4.3 Relevant predicate learning

As discussed in chapter 2, Zadeh described human cognition in terms of granulation, organization and causation (Zadeh, 1997).

The goal of Regis Pierrard's PhD was to extract knowledge from data using predicates (in particular, fuzzy relations), i.e. taking advantage of the expressiveness developed earlier (see chapter 3). Such an approach concerns the two first points of human cognition: granulation and organization.

Indeed, the idea behind Regis's PhD was to consider any object O that can be divided into granules o_1, \dots, o_n (with $n \geq 1$) and to use relations between those granules either to recognize O (classification) or to recognize a subset of the granules, not necessarily individually (annotation). Regis's PhD covered the training phase, to learn the relations automatically, the classification/annotation, and finally the explanation of such decisions. In this section, we focus on the two first stages, the explanation being developed in chapter 5, section 5.2.2. Moreover, we only considered an image as an object O , and regions or objects as the granules o_1, \dots, o_n . In Regis's manuscript, the reader can read the generalization of this idea (Pierrard, 2020).

The application on images was motivated by the fact that understanding an image relies on the comprehension of the relations between the entities (e.g. regions, objects) it contains (Biederman, 1981; Geurts, 2001). The granulation is hence performed beforehand by any object detector or segmentation method.

The next subsection discusses our inspiration from the principles of generalization and specialization.

4.3.1 Generalization versus specialization

These terms are often used in machine learning. The consequence of a good generalization is that the model will work on examples that have not been encountered during the training phase. In the contrary, machine learning often associates specialization with overfitting.

However, we are closer to the fields of inductive learning and logic. Indeed, inductive learning consists in, given a series of known positive examples and counterexamples about a concept, generalizing a concept description ("Inductive Learning" n.d.). In that field, generalization and specialization are related to two kinds of approaches:

- Top-down approaches start with the most general clause (the empty clause) and specialize (or refine) it until it no longer covers negative examples.

- Bottom-up approaches start from examples and generalize them into more global concepts ("theories"). They use generalization operators that are reversions of the classical deductive rules (unification, resolution and implication).

Finally, the main motivation remains a personal experience. With the kid of a friend of mine, we used to go to a park to observe ducks, waterhens, etc. When he was 3, we went to a zoo, and he looked at a pelican, and told me: "Look at the duck!". Of course, it has a beak and wings as ducks. It was generalization: basically, the concept of duck has been generalized regarding attributes, but that was not sufficient to discriminate a duck from a pelican. Once taught, specialization occurred and new attributes (e.g. the size, the color, the neck) were considered to distinguish both kinds of animals.

We thus built an inductive approach whose generalization is based on frequent predicates that are also suitable to define the considered classes of objects, and regarding how the model will be used.

We now describe the approach to extract relevant predicates from a dataset to perform either classification or annotation.

4.3.2 Extraction of relevant predicates

Let \mathcal{V} be the vocabulary used by the model we want to build: in the case of images, it contains spatial relations, or more generally, spatial predicates. To build a model for image classification or annotation, we proposed to extract the relevant predicates between the entities in the images. The intuition behind is that well-chosen predicates applied on some well-chosen granules would extract a sort of definition of concepts that are important to describe the image. Hence, we asked ourselves: what is a relevant predicate?

We based our work on the following postulate: *a relevant predicate is frequent in the training set*. This is justified by the fact that relevant features should occur consistently whereas irrelevant features should occur inconsistently (Kellogg, 1980). In other words, entities from one given class should share the same relevant predicates. The limits of this postulate is that learning on few instances could be highly impacted by the presence of one or several outliers.

Frequent predicates mining

As a first step, we proposed to select frequent predicates in a dataset regardless of their discriminating power. To perform classification or annotation, they were then embedded in higher-level structures that were both descriptive and discriminative (Hendricks et al., 2016; Lesot, Rifqi, and Bouchon-Meunier, 2008) (see sections 4.3.4 and 4.3.5). Since we were looking for the frequent subsets of predicates of each class, we decided to carry out the learning phase using a one-vs-all approach. The learning was thus performed class by class. We relied on fuzzy frequent itemset mining.

Frequent itemset mining aims at extracting frequent patterns in a database. It has originally been introduced for performing association rule learning (Agrawal, Imieliński, and Swami, 1993). In such problems, the goal is to build rules that catch the frequent patterns in the database. The most common example of association rule learning is the *market basket problem*. In this problem, we have a dataset of transactions made by customers. Each transaction contains items that one customer purchased. Based on this dataset, the objective is to extract rules that describe well the behavior of consumers. The vocabulary is thus mainly borrowed from this application, e.g. "items" and "transactions".

In the fuzzy counterpart, simply known as fuzzy frequent itemset mining, the relation between transactions and items is fuzzy. Nevertheless, the results are still crisp itemsets.

Formally, fuzzy frequent itemset mining is performed on fuzzy formal contexts.

Definition 4.3 (Fuzzy formal context (Belohlávek, 2012))

A fuzzy formal context is a tuple $(\mathcal{T}, \mathcal{I}, \mathcal{R})$ such as:

- \mathcal{T} is a set of transactions,
- \mathcal{I} is a set of items,
- $\mathcal{R} : \mathcal{T} \times \mathcal{I} \rightarrow [0; 1]$ is a dyadic fuzzy relation that expresses to which extent items belong to transactions.

We chose to fuzzify the *Close* algorithm (Pasquier et al., 1999). The principle is it finds all the frequent closed itemsets so that it can work on a more compact representation of the frequent itemsets. Since there are often less frequent closed itemsets than frequent itemsets, the search space is smaller, the computation is less costly and the number of database passes is reduced.

More formally, our goal was to extract the frequent subsets of predicates $I \subseteq \mathcal{I}$ such that $\text{support}(I) \geq S$. The support of a frequent subset is simply its frequency in the database. Given a closure operator, the Fuzzy Close algorithm is able to do that in two steps (Pierrard, Poli, and Hudelot, 2018a):

1. Determine all the frequent closed sets of relations. This is achieved iteratively: we first search all the frequent closed singleton of relations, then the frequent closed sets of size 2, and so on until we obtain all the frequent closed sets of relations.
2. Derive all the frequent subsets of relations from the frequent closed sets of relations.

When dealing with a dataset whose instances are highly correlated with each other, the number of closed frequent subsets of relations is much lower than the number of frequent subsets of relations. Therefore, after the first step, we can derive frequent subsets of relations from just a few frequent closed subsets. That is why, in a setting where the dataset contains correlated data, this algorithm is faster than alternatives from the literature.

At the end of this step, we have for each class a set of predicates that can be used for building rules or constraints.

The next paragraph is about obtaining a fuzzy formal context from an image dataset.

From image dataset to fuzzy formal context

The approach we proposed requires evaluating fuzzy predicates from the vocabulary \mathcal{V} before learning the most frequent among them. At the end of the evaluation step, the system has a dataset that can be represented as a formal fuzzy context on which it can perform fuzzy frequent itemset mining, as seen above.

Contrary to relational learning, we do not have the values of the different predicates in a dataset. From the granules of the training images, we automatically evaluate the predicates.

The time complexity of this step directly depends on the number of relations that are evaluated. Let us introduce the following notations:

- Let $\mathcal{V} = \{P_1, \dots, P_{n_{\mathcal{V}}}\}$ be the vocabulary as a set of $n_{\mathcal{V}}$ predicates.
- Let $\alpha : \mathcal{V} \rightarrow \mathbb{N}$ be a function such as $\alpha(P)$ denotes the arity of the predicate P for each $P \in \mathcal{V}$.
- Let \mathcal{X} be the space where instances of the dataset are defined.
- Let x be an image of the dataset.

- Let $\mathcal{O} = \{o_{x,1}, \dots, o_{x,K} \mid o_{x,i} \in x, \forall i \in \llbracket 1; K \rrbracket\}$ be a set of K granules in \mathbf{x} .
- Let $E_{\mathbf{x}}(\mathcal{V}) = \{P(o_{\mathbf{x},1}, \dots, o_{\mathbf{x},\alpha(P)}) \mid P \in \mathcal{V}, (o_{\mathbf{x},1}, \dots, o_{\mathbf{x},\alpha(P)}) \in \mathcal{P}(\mathcal{O}_{\mathbf{x}})\}$ be the set of all the relations in \mathcal{V} evaluated on the granules in $\mathcal{O}_{\mathbf{x}}$. $\mathcal{P}(\mathcal{O}_{\mathbf{x}})$ is the power set of $\mathcal{O}_{\mathbf{x}}$.

The total number of evaluations to compute on the whole training set \mathcal{D} that contains n images is:

$$|E_{\mathcal{D}}(\mathcal{V})| = \sum_{i=1}^n \sum_{j=1}^{n_{\mathcal{V}}} \frac{|\mathcal{O}_{\mathbf{x}_i}|!}{(|\mathcal{O}_{\mathbf{x}_i}| - \alpha(P_j))!} \quad (4.4)$$

In particular, this quantity directly depends on the number of relations and their arities. Moreover, some relations may be compute-intensive, which makes the whole step longer.

In the next subsection, we describe the three different heuristics we proposed to keep $|E_{\mathcal{D}}(\mathcal{V})|$ acceptable.

4.3.3 Heuristics to decrease the number of predicates to evaluate

We distinguished two kinds of optimization to the problem of the number of evaluations: local optimizations are performed relation by relation, contrary to global optimizations that allow selecting predicates to be evaluated regarding previous evaluations. Both kinds of strategies are compatible since the first kind aims at reducing the computation time of one evaluation of a specific relation while the second kind enables to prevent unnecessary computations.

We proposed a local optimization for spatial relations based on fuzzy mathematical morphology (see chapter 2, section 2.2.5 for equations). This was a collaboration with Laurent Cabaret (MICS, CentraleSupélec). The principle is to reorder the loops in the fuzzy morpho-mathematical operators and to use vectorization. Vectorization denotes a set of CPU optimizations with specific instructions like SIMD, AVX, etc. The reader is invited to read (Pierrard, Cabaret, et al., 2020) for more details.

We also proposed two heuristics as global optimizations. The first one consists in pruning the predicates that are not frequent in the dataset. We assume here that a relation that is, on average, fully satisfied in less than half of the instances in the training set is not representative of the class under study. Thus, that enables to detect relations whose current support prevents their final support to be greater than or equal to 0.5. This also presents the advantage of being independent from the vocabulary and from the task to perform.

The second heuristic directly uses the knowledge about the predicates as introduced in chapter 3, section 3.5. This knowledge enables to express links between predicates. In this work, we are interested in three kinds of links: dependency, implication and symmetry. The principle is to propagate the information of evaluated relations (using the links between relations) to gain insight on non-evaluated relations. This materializes as an order of evaluation on relations with the relations conveying more information at the front.

After the selection of relevant predicates, we proposed to build a model for classification or annotation.

4.3.4 Predicate learning for image classification

For a classification problem, a fuzzy rule-based classifier (see chapter 2 for the principles) is used. Let \mathcal{Y} be the set of all possible labels where each label $y \in \mathcal{Y}$ is associated with an image. The classification can be defined as a function \hat{f} that associates to an image one of the labels.

For each $y \in \mathcal{Y}$, we got a set MFC_y of maximal frequent closed subsets of predicates by applying the Fuzzy Close algorithm. Every subset of MFC_y are transformed into a rule.

To get discriminative rules, we remove from the subsets all the relations that are common to several classes to get the set of discriminative descriptors MFC_y^* such as:

$$MFC_y^* = \left\{ I^* \mid I^* = I \setminus \bigcup_{y' \in \mathcal{Y} \setminus \{y\}} \left(\bigcup_{J \in MFC_{y'}} J \right), I \in MFC_y \right\}. \quad (4.5)$$

Thus, we ensure that we get sets of predicates that are both descriptive and discriminative. We assumed here that the vocabulary has been set properly so that there is at least one $I^* \in MFC_y^*$ such that $I^* \neq \emptyset$.

Then, for each $y \in \mathcal{Y}$, we build a rule for each set of predicates I^* in MFC_y^* such as

$$\text{IF } \bigwedge_{P \in I^*} P \text{ THEN label} = y. \quad (4.6)$$

Since there may be several descriptors for a given class, rules are actually aggregated following a fuzzy inference process (Magdalena, 2015), so that for all image x :

$$\forall y \in \mathcal{Y}, \mu_y(x) = \bigvee_{I^* \in MFC_y^*} \left(\bigwedge_{P \in I^*} P \right) \quad (4.7)$$

with $\mu_y(x)$ the membership degree of x to the class represented by label y and \bigvee an aggregation operator, such as the supremum or the mean.

However, this definition does not rely on the support of the subsets of predicates, which brings valuable information about their reliability. Indeed, descriptors do not all have the same support and so the rules they entail should not all have the same weight in the final decision. Thus, we take into account the support of each descriptor to weight the rules as proposed in fuzzy inference systems (Magdalena, 2015), for all image x :

$$\forall y \in \mathcal{Y}, \mu_y(x) = \bigvee_{I^* \in MFC_y^*} \left(\text{support}(I^*) \times \bigwedge_{P \in I^*} P \right) \quad (4.8)$$

where we can interpret $\mu_y(x)$ as the confidence in assigning the label y to the instance x . Then, the predicted label \hat{y} is the label associated to the highest confidence:

$$\hat{y} = \underset{y \in \mathcal{Y}}{\text{argmax}} \mu_y(x).$$

4.3.5 Predicate learning for image object or region annotation

Fuzzy Constraint Satisfaction Problems (FCSPs), as defined in chapter 2, can be suitable to solve image annotation problems in which the labels and the objects to annotate are known (even if they are automatically detected, by a segmentation of the input image for instance) (Vanegas, Bloch, and Inglada, 2016). The intuition behind that approach is that such an annotation problem can be combinatorial and the labels are affected accordingly to each other, by opposition to individually like in classical approaches. The annotation of an image is a function \hat{f} that maps a subset of the granules (representing regions, objects) of the image with labels.

In such a context, the set of variables X corresponds to the objects we would like to retrieve. These variables share the same domain D that represents the regions obtained by segmentation. The constraints in C are defined by fuzzy relations between variables, which may involve groups of n objects ($n \geq 1$) (Vanegas, Bloch, and Inglada, 2016).

In our case, the generation of C is analogous to the generation of rules: it is performed label by label and it is based on extracting the set of maximal frequent closed subsets of relations using the Fuzzy Close algorithm. However, unlike rules, the descriptors from the relations that are common to several classes are not pruned. This prevents from having

too few constraints in the FCSP and we just ensure no descriptor describes several classes. The risk if too few constraints are learned is there may be too many highly consistent solutions. On the other hand, we could get no consistent solution if too many constraints are learned, but this case should not happen if the various minimum supports (for each class) are set properly.

The difference with the previous case is that we only retain the descriptor with the largest cardinality. If there are several such descriptors, we select the one that has the greatest support in the training set. Let $y \in \mathcal{Y}$ be a label. Let I_y^M be the descriptor in MFC_y with the largest cardinality such that

$$I_y^M = \operatorname{argmax}_{I \in \mathcal{J}} [\operatorname{support}(I)] \text{ such as } \mathcal{J} = \{J \in MFC_y \mid |J| = \max_{P \in MFC_y} |P|\} . \quad (4.9)$$

As explained above, we assumed the itemset of the largest cardinality will be the most helpful for solving the problem since it enables generating more constraints. Since constraints are fuzzy, we preferred having more constraints that may lead to a smaller degree of consistency than fewer constraints that may not be enough to solve the problem. In addition, the union of frequent maximal itemsets is not an acceptable choice since it is not a frequent itemset (otherwise, it would be the only one maximal frequent itemset).

We know that each evaluated predicate \mathcal{P} in I_y^M links one or several classes of entities. Let Ω_P be a set that contains those classes. In the definition of the FCSP, each variable is associated with a different class of entities. Therefore, for each item in I_y^M , we generate a constraint $(\mathcal{P}, V_{\Omega_P})$ with V_{Ω_P} the set of variables corresponding to the set of classes Ω_P .

After generating constraints for each predicate and for each class, the obtained set of constraints is used for defining and solving a FCSP. Given a new instance, the most consistent solution to this problem will lead to the annotation of every entity in the instance under study. As for rules, a *confidence* degree can be computed for a given annotation $y \in \mathcal{Y}$ as the product of the support of the descriptor I_y^M and the evaluation of the least consistent constraint in I_y^M .

Some constraints might appear several times because several different classes produced the same constraints. That happens with symmetrical p -ary relations with $p > 1$. In that case, the set of constraints is reduced so that it contains this constraint only once.

Finally, in case of a missing granule (e.g. object, region) in the image, as an anomaly, we can distinguish two cases:

- The missing granule is involved in all the constraints: in this case, the FCSP has no solution and there is no way to perform the annotation. This is the worst case but it has some sense not to perform annotation if the central object/region is missing.
- The missing granule is involved only in a few constraints. Then the FCSP will have some solutions, but the overall satisfaction, thus the confidence, will be lower than without this anomaly.

4.3.6 Application to medical image annotation

We experimented the approach on real images during the DeepHealth project. We used a public dataset that comes from the VISual Concept Extraction challenge in RAdioLogY (VISCERAL) project (Langs et al., 2013). This project proposed a cloud-based infrastructure for the evaluation of medical image analysis techniques in Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). The project exposed one dataset called *Anatomy3* (Jimenez-del-Toro et al., 2016). This dataset was used in one segmentation benchmark. It is composed of 391 CT and MRI images.

In these images, up to 20 different organs are segmented. Segmentation files are provided as binary images for each organ. Thus, the granules we deal with are not fuzzy.

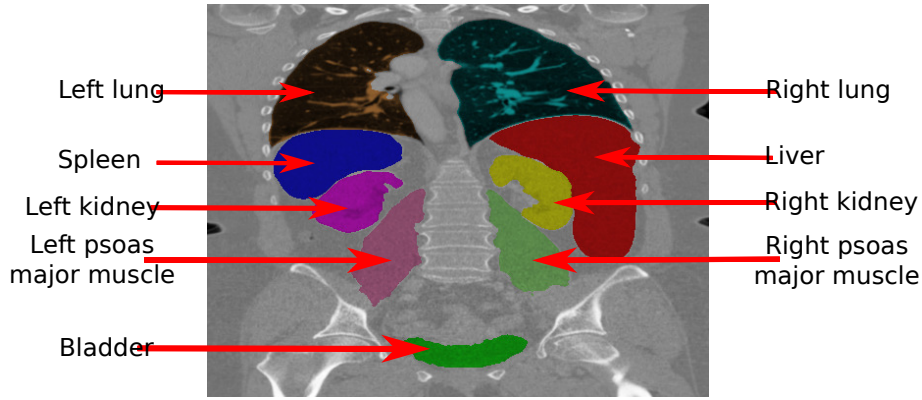


Figure 4.7: The nine organs of interest in this experiment (in color).

From these images, we created our own dataset for the purpose of this experiment. Since most images from the original dataset contain few segmented organs, we selected the instances containing the 9 following organs: the liver, the spleen, the urinary bladder, the left and right kidneys, the left and right lungs and the left and right psoas major muscles (see figure 4.7). In the end, we had 35 images and 315 regions. The goal was to label them in each instance. This new dataset was small, which enabled us to assess how our model can perform by learning from few examples.

The vocabulary \mathcal{V} was limited to 9 spatial predicates whose relations are shown in figure 3.26. Applying the third heuristic, we obtained the following order of evaluation:

- | | |
|-----------------------------|------------------------------------|
| 1. <i>IsSymmetricalTo</i> ; | 6. <i>ToTheLeftOf</i> ; |
| 2. <i>Above</i> ; | 7. <i>ToTheRightOf</i> ; |
| 3. <i>CompletelyAbove</i> ; | 8. <i>CompletelyToTheLeftOf</i> ; |
| 4. <i>Below</i> ; | 9. <i>CompletelyToTheRightOf</i> . |
| 5. <i>CompletelyBelow</i> ; | |

For semantic image annotation, we proposed the following workflow:

1. All the predicates from the vocabulary were assessed on the granules of each instance in the training set. Predicates were evaluated according to the order given by the third heuristic. The third heuristic helped discarding infrequent relations online.
2. The most frequent subsets of predicates among the ones that have been assessed in step 1 are extracted. To perform this task, we split the training set according to the class (organs). We then applied the fuzzy Close algorithm for each subset. Thus, we obtained one or several sets of relevant relations for each class.
3. We generated the corresponding FCSP:

- The set of variables:

$$X = \{x_{\text{liver}}, x_{\text{spleen}}, x_{\text{bladder}}, x_{\text{r_psoas}}, x_{\text{l_psoas}}, x_{\text{r_lung}}, x_{\text{l_lung}}, x_{\text{r_kidney}}, x_{\text{l_kidney}}\}.$$

- The domains were the same for each variable and contain all the segmented regions of the image.
- The flexible constraints were generated from the frequent subsets of relations class by class as explained before. We got a set of constraints C . Furthermore, since every organ is unique, there cannot be identical labels in this problem. That means C had to be extended with the *AllDifferent* global constraint.

Then, once the FCSP was defined, for a given example, it was solved using the FAC-3 algorithm, for filtering inconsistent domain values, and the backtracking algorithm, for exploring the possible solutions. In the end, the entities of interest had been labeled and an explanation was produced based on the constraints that were derived from the frequent itemsets (see chapter 5, section 5.2.2).

We investigated the number of training examples needed for our model to perform well. Using the nested cross-validation, we evaluated the performance of the model for a number of training examples ranging from 17 to 34. Then, performing a reverse cross-validation (the training set and the test set were inverted), we assessed the performance of the model for a number of training examples ranging from 1 to 17. In that situation, instances were part of the test set in several iterations, but we ensured that the model had learned from all the possible combinations of instances. From 7 to 34 training examples, the model reached an accuracy of 100%. Then, we got 99.6% for 5 training examples, and 99% for 3 and 2 training examples. These results show that the model can learn valuable information from a small dataset. In this experiment, the whole dataset did not contain any outlier (like a missing organ for example) so it is suited to learning from few data.

Before training, 21420 relations had been evaluated over the whole dataset. We evaluated our first heuristic on these evaluations. Overall, this strategy enabled to prevent 32% of all the evaluations. The second heuristic enabled us to avoid computing 7.6% of the total number of evaluations. Overall, by combining both heuristics, we managed to avoid computing about 40% of the total number of evaluations (8596 out of 21420 evaluations). In particular, it enabled us to avoid computing expensive relations like symmetry or morphological directional relations.

4.4 Extraction of gradual rules from experimental data

4.4.1 Context

This work is part of a collection of projects that consist in developing AI methods for materials science, spurred on by Frédéric Schuster, in charge of CEA’s programme on materials and processes.

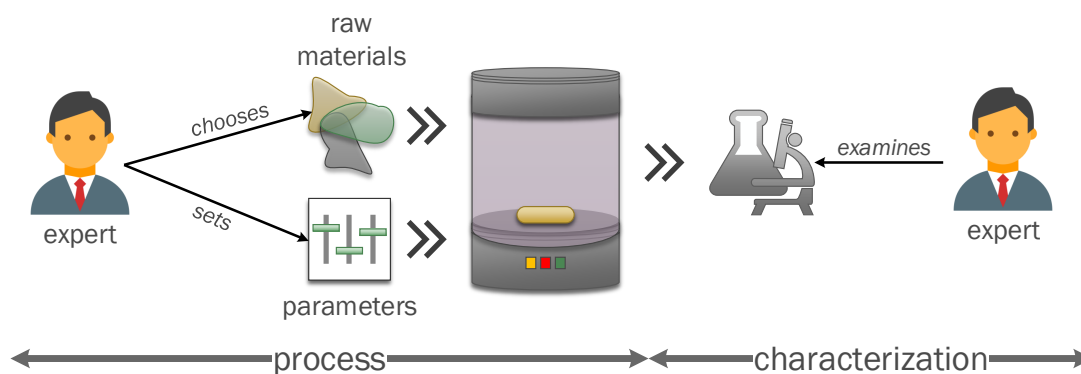


Figure 4.8: Overview of materials science (Poli, Hajri, and Boudet, 2021).

Materials science is a research field that focuses on the design and discovery of new materials. Advances in materials science are related to the identification of relationships between the process by which a material is produced, the microstructure or the nanostructure of the material, its properties and its performance. Once uncovered, these relations can guide the design of new materials (Poli, Hajri, and Boudet, 2021). Each production process has its own parameters and each product has its own properties (e.g. mechanical, physical). The values of these properties determine the performance of the product. Thus,

it is important to detect which parameters have the most important impacts on the product properties and then on its performance. For this reason, we proposed to mine causal relations between processing parameters and materials properties.

AI approaches to material design do not need a precise mathematical model of the process, compared to simulations. Figure 4.8 presents my general conception of material design that consists of two phases. The first phase is the process and the preparation of this process by an expert. A process can be seen as a black box with parameters and inputs: in the case of material design, the inputs are the different raw materials that will be combined during the process. The result of the process is a new material. It leads to a second phase, which concerns its characterization to determine if it fits the expert's expectations. For now, we are not involved in the characterization of the materials.

During my discussion with experts in that field, I realized the importance of gradual rules in their knowledge: for instance, "the higher the temperature, the stronger the material". The idea was to develop a proof of concept to demonstrate that it was possible to predict a property of a material from its manufacturing parameters, with a strong focus on gradual rules. To tackle the whole problem, I imagined the following workflow:

1. Find which parameters have an effect on which properties. This was the occasion to investigate the extraction of causal links. To Zadeh (Zadeh, 1997), causation is the third component of human intelligence.
2. Characterize the causal links. In this case, we reject all the causal links that are not gradual.
3. Represent and parameterize a gradual rule.

It is important to notice that in this work, we do not know the nature of the relation between the parameters and a property: is the relation linear? Is it polynomial? If this relation was known, the problem would be a simple regression with the right mathematical model. However, we do not want to make any assumptions of this nature.

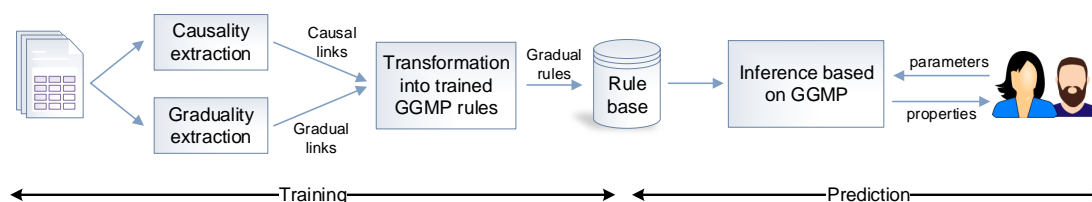


Figure 4.9: Overview of the proposed approach for gradual rule extraction.

This was the topic of Hiba Hajri's postdoctoral fellowship. We proposed an approach that is summarized in figure 4.9. During the training phase, from experimental data, we extract causality links. We also extract the gradual links. If we have causal links and gradual links involving the same variables, we can create gradual rules. During the prediction, the inference is performed and from the parameters, we are able to predict the properties of the materials (in the gradual case only).

4.4.2 Causality and graduality extraction

In the literature, several approaches use statistical analysis for defining the significance of the manufacturing processing parameters. We chose the methods that are the most used in materials science:

- In the case of categorical variables, an Analysis of Variance (ANOVA) when its assumptions of normality, homogeneity of variance and independence of observations

are met. A non-parametric statistical test, such as the Kruskal-Wallis test, is used otherwise.

- In the case of continuous variables, we perform a simple regression analysis.

For its various advantages, we used the GRAdual rANKing (GRAANK) to find the gradual itemsets in our datasets (Laurent, Lesot, and Rifqi, 2009). GRAANK combines different existing approaches to benefit from both semantic quality and computational efficiency. The authors use the Kendall's tau ranking correlation coefficient defined as the proportion of discordant pairs, i.e. the frequency of pair-wise inversions. A concordance matrix is initially built for each gradual item from all the attributes. Then, candidate gradual itemsets are generated and their concordance matrix is computed as the logical AND of the concordance matrices of the joined itemsets. A candidate gradual itemset is evaluated based on its support measure obtained with the sum of the matrix elements, divided by the total number of pairs of objects. We now present some basic notations and definitions as defined in (Laurent, Lesot, and Rifqi, 2009).

Definition 4.4 (Gradual item)

A gradual item A^* is defined as a pair of an attribute A associated to a variation $* \in \{\geq, \leq\}$. A^{\geq} expresses an increase in A values "the higher A ". A^{\leq} expresses a decrease in A values "the lower A ".

For instance, the gradual item $Price^{\geq}$ is interpreted as "the higher the price".

Definition 4.5 (Gradual itemset)

A gradual itemset $GM = A_1^* A_2^* \dots A_n^*$ is a combination of n gradual items. It implies a simultaneous change between n attributes.

For example, the gradual itemset "the higher the quality of the product and the higher its price" can be formalized by the two items: $Quality^{\geq} Price^{\geq}$.

Definition 4.6 (Gradual rule)

A gradual rule GR noted as: $GM_1 \rightarrow GM_2$, is defined as a pair of gradual itemsets GM_1 and GM_2 that have to be related by a causality link. GM_1 is the antecedent of the rule GR and GM_2 represents its consequent. This causality constraint makes the difference between a gradual itemset and a gradual rule.

For example, the gradual rule, "the faster the car, then the greater the fuel consumption" can be denoted $Speed^{\geq} \rightarrow Consumption^{\geq}$.

4.4.3 Representation and evaluation of gradual rules

Graduality has been defined in different ways in fuzzy logic (Di Jorio, Laurent, and Teisseire, 2008; Dubois and Prade, 1992; Marsala and Rifqi, 2017; Vo, Detyniecki, and Bouchon-Meunier, 2013). In this work, we first chose the Gradual Generalized Modus Ponens (GGMP) (Vo, Detyniecki, and Bouchon-Meunier, 2013) as a way to approximate monotonous functions, and thus gradual relations between parameters and properties.

Gradual Generalized Modus Ponens

GGMP is an extended version of the Generalized Modus Ponens (GMP) (see chapter 2, section 2.4.3) that allows integrating the gradual hypothesis when a monotonic relationship

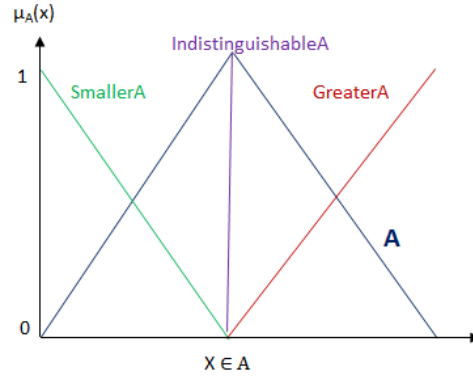


Figure 4.10: GGMP decomposition of a triangular fuzzy set A .

exists between the input and the output. The GGMP helps us to cover different scenarios:

- The user sets the process parameters in a precise way (e.g. "the temperature is 200°C"). In this case the answer is a single value for each property.
- The user sets an imprecise value, represented for instance by a fuzzy number (e.g. "the temperature is approximately 200°C"). In this case, the result is a fuzzy set.

To enforce the graduality, the GGMP consists in decomposing the fuzzy sets of the premise and the conclusion of each fuzzy rule into three parts: Smaller, Greater and Indistinguishable, as shown in figure 4.10.

For example, for an increasing GGMP, let V and W be two linguistic variables defined on the universes of discourse \mathcal{U}_V and \mathcal{U}_W respectively. A is a fuzzy set of V and B a fuzzy set of W . Let r be a fuzzy rule "if V is A then W is B ". The universe of the premise (resp. of the consequence) is decomposed into three fuzzy sets regarding the fuzzy set A (resp. the fuzzy set B): *SmallerA*, *GreaterA* and *IndistinguishableA* (resp. *SmallerB*, *GreaterB* and *IndistinguishableB*). Using this decomposition, the inference is focused on the corresponding parts of A and B . Thus, to infer the *SmallerB* (respectively *GreaterB* and *IndistinguishableB*) part of B , the *SmallerA* (respectively *GreaterA* and *IndistinguishableA*) part of A is only used (Vo, Detyniecki, and Bouchon-Meunier, 2013).

In GGMP, A and B have to be convex, normalized and continuous, and their supports have to be bounded. The definition of the membership functions of the three parts, *Smaller*, *Greater* and *Indistinguishable* of both the premise and the consequence of r , is based on the core and the complement of A and B (Vo, Detyniecki, and Bouchon-Meunier, 2013). GGMP has been also defined for rules with multiple inputs and monotonic decreasing relations (Vo and Detyniecki, 2013; Vo, Detyniecki, and Bouchon-Meunier, 2013).

The universe of discourse has to be partitioned: the level of granularity of each linguistic variable (i.e. the number of fuzzy sets partitioning the universes of discourse) included in the rule has to be chosen. It has to be set according to both the number of data and the complexity of the shape of the function to approximate. We assume that at least one triangular fuzzy set is required for a monotonic linear curve. Furthermore, in order to preserve the membership functions requirements defined in the GGMP mechanism, two half-triangles have to be added in the final partitioning, at both ends of the universe of discourse.

In the case of several inputs, the GGMP considers that the inputs contribute to the same extent to the output value. In practice, in particular in materials science, this assumption is often false. Indeed, an output value can increase faster regarding an input than another. As a proof of concept, we decided to weigh the rules to consider this difference in contribution.

Training

For the proof of concept, we chose to use a genetic algorithm to optimize the parameters of the gradual rules. We consider different sets of parameters:

- The location of the critical points of the triangular membership functions of the rules premises and consequences. Since we have a strong partition, we keep only the triangles' tops.
- The rules weights. These parameters are critical and must be constrained since some weights configuration can cancel the gradual property of the GGMP.

We used the Root Mean Square Error (RMSE) value between the predicted values and the actual ones as the fitness function.

Inference

Let r be a fuzzy rule “if V is A then W is B ” and A' be a fuzzy set that is a new observation for V . The conclusion's membership function $\mu_{B'}$ is defined as follows (Vo, Detyniecki, and Bouchon-Meunier, 2013):

$$\forall y \in \mathcal{U}_W, \mu_{B'}(y) = \sup_{x \in \psi(y)} \top(\mu_{A'}(x), \mu_{\mathcal{I}}(x, y)) \quad (4.10)$$

where \mathcal{I} is a fuzzy implication whose membership function is $\mu_{\mathcal{I}}$, and

$$\begin{aligned} \psi(y) = \{x \in \mathcal{U}_V \mid \mu_{P_A}(x) = \mu_{P_B}(y) \text{ and } \mu_{P_B}(y) > 0, \\ P \in \{\textit{Smaller}, \textit{Greater}, \textit{Indistinguishable}\}\} \end{aligned} \quad (4.11)$$

where μ_P are the membership functions of the smaller, greater and indistinguishable fuzzy sets automatically created from the mentioned fuzzy set (as in figure 4.10).

In the context of our proof of concept, considering that our datasets have only crisp data, we adapted the equation proposed in (Vo and Detyniecki, 2013) to obtain the aggregated and defuzzified value y_f :

$$y_f = \frac{\sum_{i=1}^m \sum_{j=1}^n w_i^j \mu_{B_i}(y_i^j) \mu_{B_i'^j}(y_i^j) y_i^j}{\sum_{i=1}^m \sum_{j=1}^n w_i^j \mu_{B_i}(y_i^j) \mu_{B_i'^j}(y_i^j)}. \quad (4.12)$$

where each $w_i^j \in [0; 1]$ represents the weight of the input j in the rule i ; y_i^j represents the points at which $B_i'^j$ has positive membership values. This equation is analogous to a conjunctive aggregation and the center of gravity defuzzification.

4.4.4 Validation on a toy dataset

Before applying to real world data, we wanted to show that a properly set GGMP could be used to approximate monotonous functions. We thus tested some toy monotonic functions. For each case, we generated a dataset with some sampled points in the universe of discourse of the input variable for the training phase:

- A quadratic increasing function $f(x) = 2x^2$ on $[1; 100]$. To get a Mean Absolute Percentage of Errors (MAPE) value of about 9%, 6 fuzzy sets are necessary for both x and $f(x)$. Results are shown on figure 4.11a.
- A sigmoid function $f(x) = \frac{1}{(1+e^{-x})}$ on $[-5; 5]$. Given the complexity of this relation, 7 fuzzy sets are used for both x and $f(x)$ to get a MAPE value of about 6.5%. Results are shown in figure 4.11b.

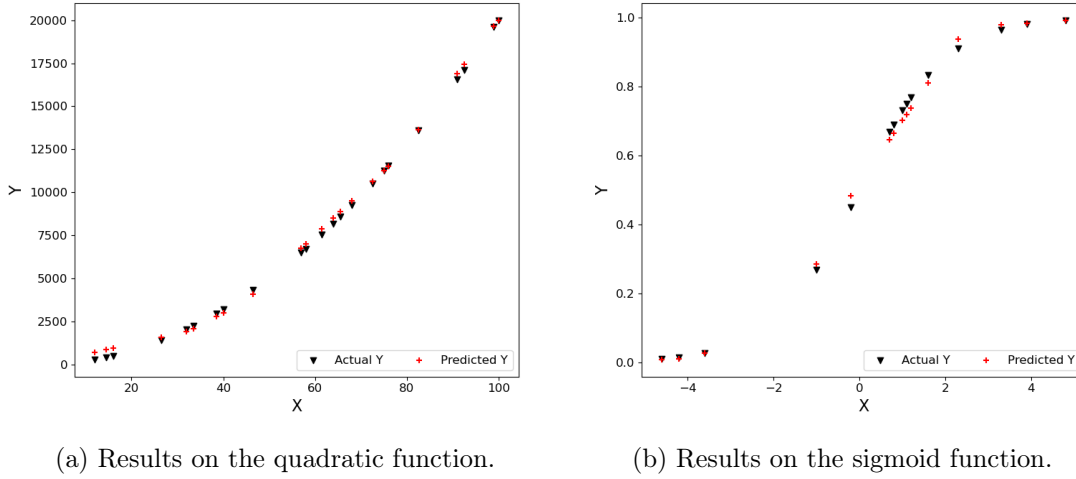


Figure 4.11: Results on a toy dataset.

4.4.5 Application to materials properties prediction

Once validated on artificial data, we applied the approach to a real-world dataset in the domain of Physical Vapor Deposition manufacturing process. The dataset was provided from a case study aiming at producing thin films of zinc oxide by cathodic sputtering. It contained 59 experiments obtained from four experiments. There were five controlled process parameters. Four of them were categorical and numerical: Pressure (Pr), Partial Pressure (PP), Power (P) and Scroll Speed (SS). Indeed, on the machine that had been used, it was only possible to select some values for these parameters. The last one, the Number of Passages (NP), was quantitative. Different types of material properties were measured: mechanical, optical and physical. Two properties will be cited after:

- the Deposition Speed (DeS) whose values ranged between 2 nm/min and 45 nm/min;
- the film's Thickness (Th) whose values varied between 7 nm and 743 nm.

The statistics tests (for both graduality and causality) indicated that DeS was more influenced by two parameters: Pr and P . We detected also two gradual itemsets using the GRAANK algorithm. Thus, we detected the following gradual link for the DeS (noted GL_1): $P \geq, Pr \leq \rightarrow DeS \geq$. We also deduced that Th is influenced by three parameters: SS , P and NP . We detected three gradual itemsets. Thus, we could conclude this gradual link holds for the Th (noted GL_2): $P \geq, NP \geq, SS \leq \rightarrow Th \geq$.

For each property, we extracted the gradual rules from the experimental data as presented before, with five fuzzy sets for each linguistic variable representing the antecedent and the consequent of each rule. We used 5-fold cross validation to evaluate the performance of our approach.

According to the results, the prediction of the property DeS has been performed with an average MAPE of 9.5% (with a standard deviation of 2.23%) and an average RMSE of 1.8 (with a standard deviation of 0.31). The prediction of the property Th has been obtained with an average MAPE of 57.49% (with a standard deviation of 5.77%) and an average RMSE of 59.12 (with a standard deviation of 19.51).

We then compared the predictive performance of our approach with the performances of other predictors from the literature:

- Adaptive Neuro-Fuzzy Inference System (ANFIS). For each input, we trained a model by testing several Gaussian membership functions ranging from 3 to 6 and respecting fuzzy strong partitioning.
- Polynomial regression with several polynomial degrees ranging from 2 to 7.

- Support Vector Regression (SVR) by testing a linear and a radial basis function kernel with default hyperparameters.
- Ensemble methods, the random forest and the Extreme Gradient Boosting (XGBoost) with monotonicity enforced.

The hyperparameters were chosen to optimize the mean and standard deviation of the metrics RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage Error).

We applied a 5-fold cross-validation using the same folds to train and test the different models. We used only singleton values for inputs, expecting singleton values for outputs, since our method is the only one to deal with more than singletons. We calculated the RMSE and MAPE metrics to evaluate the trained models for each fold. Then we considered the average and the standard deviation of the RMSE and MAPE obtained for all the folds to assess the performances of a predictor to another one.

Table 4.6 presents the results of the different models trained using the selected methods and our method to predict both the *DeS* and *Th* properties. For *DeS*, our method (adapted GGMP) gave the best results in terms of mean and standard deviation of RMSE and MAPE metrics. For *Th*, it gave the best results in terms of mean and standard deviation of RMSE. However, XGBOOST had the best performance in terms of the average MAPE.

	<i>DeS</i>		<i>Th</i>	
	RMSE	MAPE (%)	RMSE	MAPE (%)
SVR	3.61 ± 1.09	18.75 ± 7.67	108.83 ± 29.09	118.31 ± 43.81
Polynomial Regression	1.9 ± 0.5	11.83 ± 3.02	94.19 ± 61.05	43.31 ± 22.67
Random Forest	2.83 ± 1.54	12.22 ± 4.15	73.5 ± 20.98	80.12 ± 38.48
ANFIS	4.6 ± 3.9	18.10 ± 15.19	96.5 ± 25.87	75.8 ± 32.01
XGBOOST	2.62 ± 1.17	14.65 ± 9.45	59.33 ± 23.23	39.47 ± 9.42
Adapted GGMP	1.8 ± 0.31	9.5 ± 2.23	59.12 ± 19.51	57.49 ± 5.77

Table 4.6: Evaluation of the prediction of *DeS* values based on *Pr* and *P*, and of *Th* values based on *P*, *NP* and *SS*, using the different selected predictors.

The proof of concept was thus successful. We has also checked the gradual rules extracted from the data with materials scientists and they confirmed both the interest in our approach and the results we obtained.

The proposed approach allows extracting useful knowledge from a few data. On the one hand, other methods, such as ANFIS, require large datasets and their interpretability is not always guaranteed due to the number of the generated rules (Salleh, Talpur, and Hussain, 2017). On the other hand, Sugeno approaches need to know a priori the shape of the function to approximate.

The advantages of our method are:

- it is a good predictor for properties from process parameters;
- it is directly representable with a sentence, e.g. "the more ... the more ...", which is understandable by the end-user;
- it deals with singleton values as well as fuzzy sets that can represent imprecise values or intervals.

Indirectly, the gradual rules can help during the search for optimal process parameters to get some specific properties. For instance, with a rule like "the higher the temperature,

the more solid the material", it is obvious that to obtain the most solid material, the temperature must be set to its maximal value.

Nevertheless, the number of fuzzy sets to partition the domains of variables has to be chosen carefully. As I will present in the conclusion of the document, we are currently working on it during the internship of Killian Susini. Each drawback is being addressed: the causality extraction, the gradual rule modeling, the number of fuzzy sets and the genetic algorithm. We also extended the use cases, considering few or big data, as well as noisy data.

4.5 Summary

In this chapter, I presented three kinds of contributions that aim at extracting knowledge from data to build interpretable and explainable models.

The first contribution allows building automatically new features from existing ones, by combining them into mathematically sound equations. These new features are thus interpretable and their use can help transparent machine learning models to reach better performances. The idea behind was indeed to compensate for their lack of complex representation of the feature space compared to deep neural networks. Without loss of generality, in this manuscript, we embedded the feature construction into decision trees induction algorithms and into FURIA rule induction. In Noëlie's manuscript, it has also been achieved for Generalized Additive Models. We also proposed a way to evaluate the interpretability of such features with a survey questionnaire.

The second contribution allows for solving problems like image classification or object (or region) annotation. It generates a transparent model in both cases, respectively a rule base or a fuzzy constraint satisfaction problem. Both models are populated by rules or constraints that are built upon a vocabulary of predicates. The principle of this approach is to extract the relevant predicates that help defining the target concepts (e.g. classes or labels). This method relies on the expressiveness that has been developed in chapter 3. The models are thus interpretable but they are also explainable since it is possible to generate a textual explanation of the decision. The next chapter, in section 5.2, develops the way textual explanations are generated in both cases.

Finally, we ended this chapter by presenting a successful proof of concept. The goal was to be able to represent and automatically extract fuzzy gradual rules to approximate monotonous functions. We recently started exploring the causality aspect, which is an important aspect of human capabilities. This is typically a high-level knowledge that can be not only used directly by researchers as insights from their experimental data, but also to perform some regression. The results are satisfying and, as I am writing this document, we are finishing the work.

Chapter 5

Improving usability and acceptance of fuzzy systems

5.1	Touch graphical user interface for rule authoring	101
5.1.1	Vocabulary edition	103
5.1.2	Rule edition	103
5.1.3	Evaluation	104
5.2	Generation of textual explanations of fuzzy systems decisions	109
5.2.1	First case study: fuzzy classifier	109
5.2.2	Second case study: semantic annotation	114
5.2.3	Human-based evaluation of explanations	117
5.3	Summary	120

As said before, one of the goals of our research is to let end-users, who are not familiar with Artificial Intelligence and logics, handle a XAI system. In this chapter, we pay attention to both the usability and the acceptance of fuzzy systems. The contributions to raising fuzzy systems to the level of XAI are twofold.

First, we observed that most of the expert system shells suffer from a lack of simplicity for authoring rules, or more generally knowledge. It is often based on a specific syntax that may be difficult to handle by end-users. We studied the cognitive impact of rule authoring and proposed a new ergonomic graphical user interface to reinforce the usability of fuzzy systems.

Then, the second aspect concerns the reinforcement of trustworthiness, that is to say, acceptance. It is helpful, especially when end-users have to make decisions according to an automatic tool. We thus focused our work on the generation of textual explanations of decisions of fuzzy systems.

5.1 Touch graphical user interface for rule authoring

There exist different methods to author rules that could be adapted to the particular case of fuzzy logic. Figure 5.1 shows some examples of graphical user interfaces (GUI) that have been developed by companies. They are generally a compromise between how the system will interpret these rules and the freedom given to the user during the authoring of the rules. Few of them take advantage of touch screens that have become a natural interaction for the youngest to oldest users. A dilemma remains unresolved: moving away from a

textual representation makes the users confused whereas text capture is not practical on touch devices.

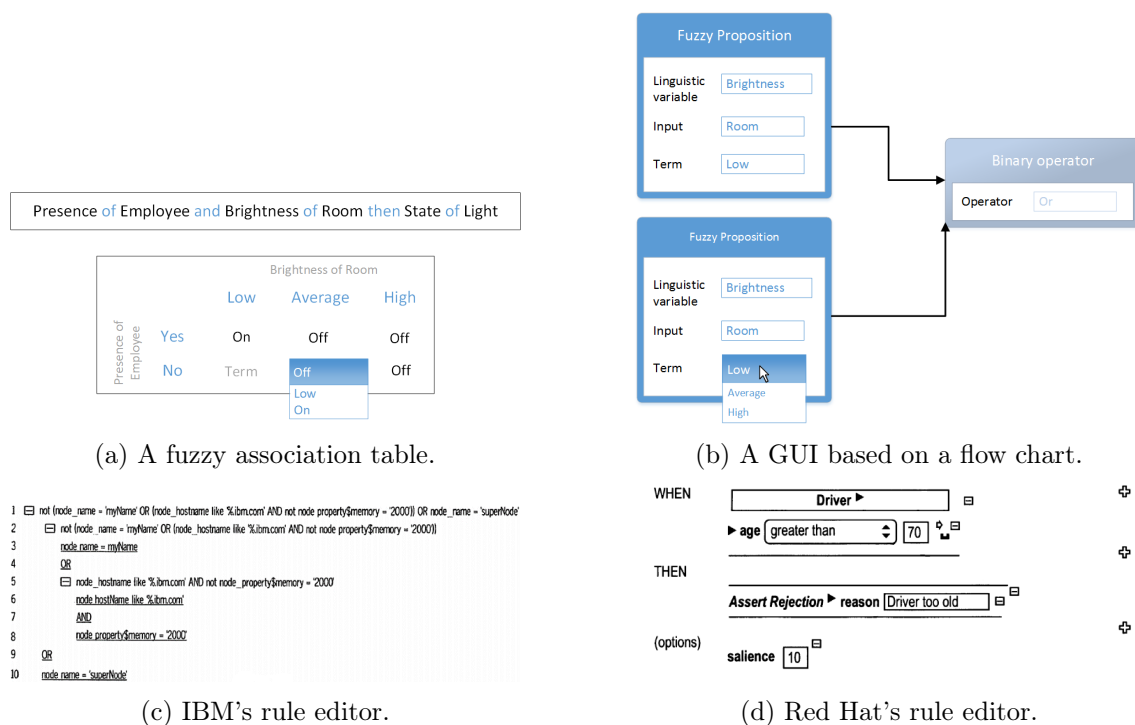


Figure 5.1: Examples of GUI to edit rules.

Our approach relied on our industrial partners, the users of ExpressIF[®], studying their needs. Globally, they wanted a modern GUI, with a focus on mobility, less mathematical or logic. Regarding the state-of-the-art, our contribution was about new devices with capacitive touchscreens. *Drag-and-drop* on such screens are used to target an audience of different profiles and ages. We also wanted to keep a form close to natural language while offering functionality similar to the existing interfaces.

The main difficulty came from the nature of expert systems. They consist of a generic inference engine, but domain-specific rules. Thus, they can be applied to different domains and the rules can be authored by different kinds of users. Contrary to specific interfaces, the rule editor had to be as generic as possible.

The study of these ergonomic issues started with the internship of Jean-Paul Laurent during his master 2 in ergonomics, in 2013. To the best of our knowledge, this was the first time such a study about ergonomics and fuzzy systems has been carried out.

The interface was designed such that the first screen allows editing the input and output linguistic variables, and eventually crisp inputs and outputs. Once the vocabulary set, the second screen allows completing the rule base with a "filling in the blanks" design so that all users can easily and quickly produce rules. The blanks form a pattern that we call a "ghost", in reference to the pale color it takes as long as it is not completed. It is possible to choose a general form of rule, then the types of propositions that compose the rule. It is also possible to select and move the different parts of a rule, or insert a new operator. The insertion of a new operator may lead to the insertion of a ghost that describes the form that should take the missing part of the current expression. All these operations are achieved by *drag-and-drop*. Figure 5.2 presents a view of the GUI with a rule being edited, in which appears the ghost of a proposition.

At this stage, the GUI would be too permissive allowing to move operators or expressions anywhere. We have adopted various measures to ensure that the user is guided while composing rules: *feedforwards* show possible actions and *feedbacks* help to focus on the

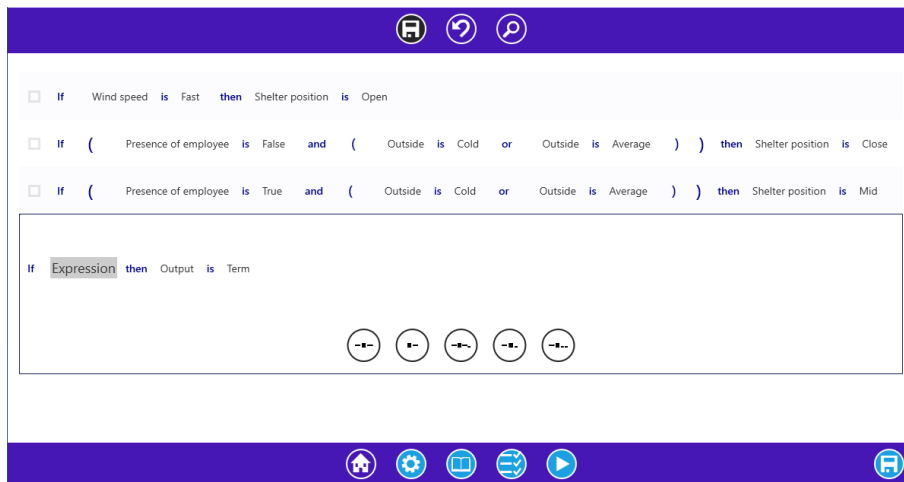


Figure 5.2: Overview of ExpressIF[®] Rule editor.

result of the ongoing action.

5.1.1 Vocabulary edition

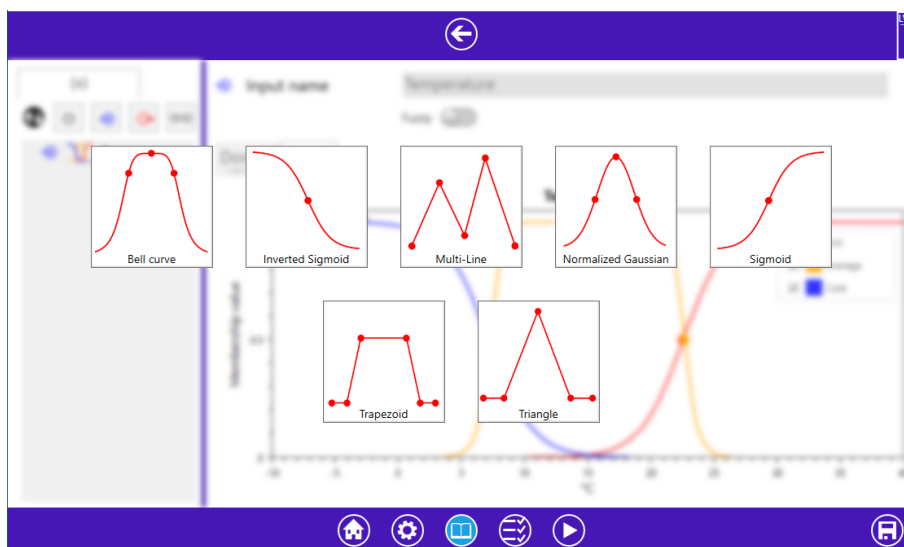


Figure 5.3: The choice of shapes for a membership function.

Once the domain has been chosen, the user can select the shape of the curve to be inserted (figure 5.3). The curve is inserted in the middle of the domain. When the curve is selected, handles appear: the user has just to move the handles to change the parameters of the curve (figure 5.4). In the case of a multiline curve, a click can add a point and break the selected line. Some users asked us to add a way to parameterize precisely a curve. A click on a handle makes a box appear to type the value of the parameters or the coordinates of the point for a trapezoid, triangular or multiline membership function.

5.1.2 Rule edition

To create a rule, the user has just to click on the button "Add" materialized by a "+" symbol. The rule being edited is highlighted in a box while leaving other rules visible. The rule is then materialized by a textual pattern containing a ghost of the premise proposition and the conclusion. Clicking the premise ghost makes a pie menu appear to select the

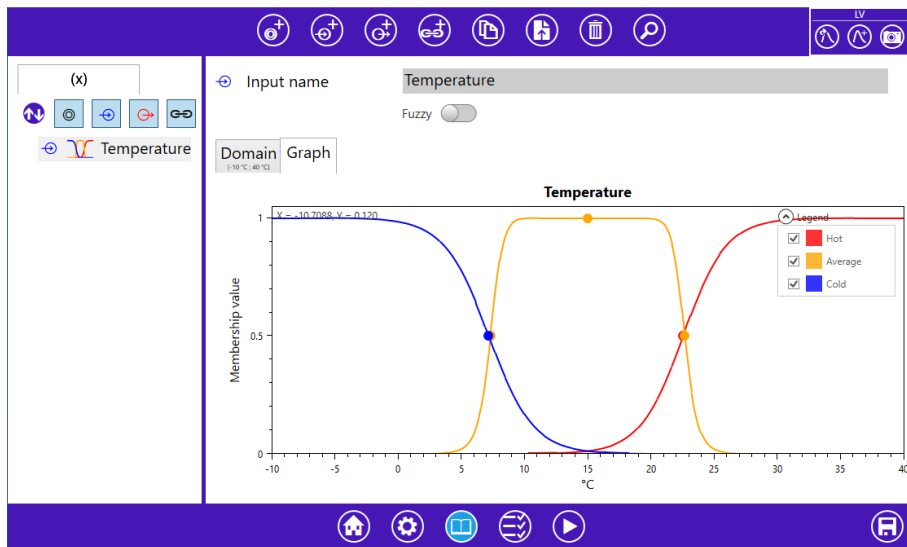


Figure 5.4: Example of a linguistic variable and its terms. The points are the handles to move.

desired type of proposition, either in the form `linguistic variable input IS term` or the form `input EQUALS value`. Each customizable term (in lowercase in the examples above) can be changed with the help of a list of options. Figure 5.2 shows a rule pattern with its various ghosts "Proposition", "Magnitude", "Output" and "Term".

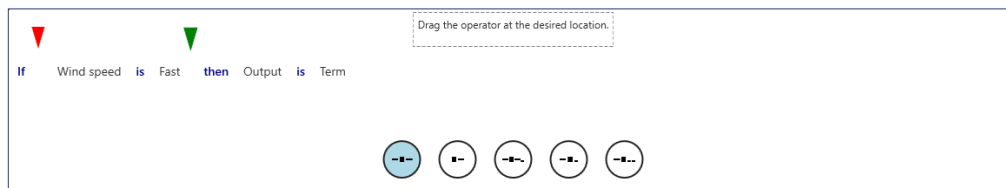


Figure 5.5: Focus on the possible destinations while *drag-and-drop* of the binary operator.

The user can create an expression by dragging its icon to a portion of the premise of the rule. Once the icon is dropped in the correct place, a ghost of the operator is placed, eventually followed by a ghost of an expression if the operator needs more operands. Animated red arrows indicate where the operator may be dropped during the drag phase (Figure 5.5).

Clicking the ghost "Op." of the operator makes a pie menu appear that shows the available operators (Figure 5.6). After the selection, the ghost is replaced with the selected operator. To change the operator, the user can simply click on it again to call back the pie menu, in which all operators are represented by icons. If a mouse is detected and it hovers the icon, it displays a context-sensitive help message. If there is no mouse, a long press on the icon makes the tooltip appear.

In the following section, we describe the evaluation of this interface.

5.1.3 Evaluation

This work falls within the framework of ergonomic studies and we therefore wished to use a panel of users to evaluate the GUI. We first describe the experimental protocol and the panel of users.

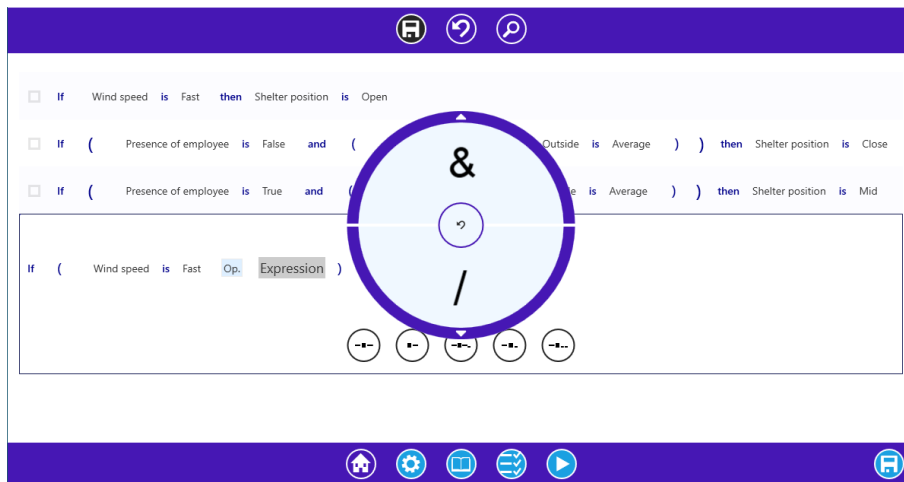


Figure 5.6: Pie menu for binary operator selection.

Protocol

Tools for editing rules have already been compared in the literature: flowcharts versus association matrices (Subramanian et al., 1992), structured language based on the English language versus flowcharts and matrices (Vessey and Weber, 1986).

It is still difficult to have access to ready-to-use GUI for fuzzy expert systems. We thus compare our GUI (referred to as *Rule editor*) to the most common tools: *Rule matrix* (see figure 5.1a) and *Rule flowchart* (see figure 5.1b). To have a look-and-feel similar between the three GUIs, we implemented our own versions of the two concurrent tools.

To compare the three tools, we used the following criteria:

- the *adoption time* measures the time required for the user to become familiar with the tool;
- the *average time for rule edition* allows measuring the efficiency of the tools;
- the *number of erroneous rules* weights the previous criterion, to avoid editing rules very fast and having only wrong rules;
- A *questionnaire* allows getting verbatim answers.

In our various applications, rules can be edited by different profiles of users: end-users, experts in fuzzy logic, scientists who are not specialists in expert systems and fuzzy logic, etc. We thus tried to have a panel that reflected this variety of users: it consisted of 27 varied users (14 men and 13 women), ages between 20 and 57. In terms of profiles, we selected 6 expert users (who use fuzzy expert systems in their business), 8 scientists (who are not familiar with fuzzy expert systems) and 13 non-scientific users. No users were the partners we interviewed before.

We chose a simple topic for tests: home automation. Users were invited to write rules to control automatic shutters according to various observed criteria: the room occupancy, the brightness, the wind and the indoor and outdoor temperatures. The test was divided into three exercises that must be solved with each of the three interfaces:

- The first exercise consisted in typing rules based on two inputs only. No help was given to the users about the use of the three interfaces. The goal was to master the different tools. The time of completion of this exercise reflected partly the adoption time.

- Between the first and the second exercise, the users who failed to master at least one interface received explanations. Exercise 2 also involved writing rules based on two inputs. This time, the execution time reflected how quickly it was possible to write a rule.
- The third exercise consisted in writing more complex rules that cannot be authored with the association matrices. The users were asked to account for this limitation alone.

In practice, the tests were quite long (between one and two hours per user). It explains partially why the panel remained restricted.

In order not to introduce biases in the assessment, two measures were taken. Firstly, the instructions about the rules to write were given using causal constructions different from "if...then...", e.g. "when the wind is strong, the shutters should be raised". This allowed not considering the difficulty of the user to carry out logical reasoning. Secondly, it was decided to present the GUIs in random order so that no technique is discriminated.

Participants were asked to complete a very detailed questionnaire in order to identify factors that may explain their performance, but statistical tests showed that only the level of education had an impact.

Results

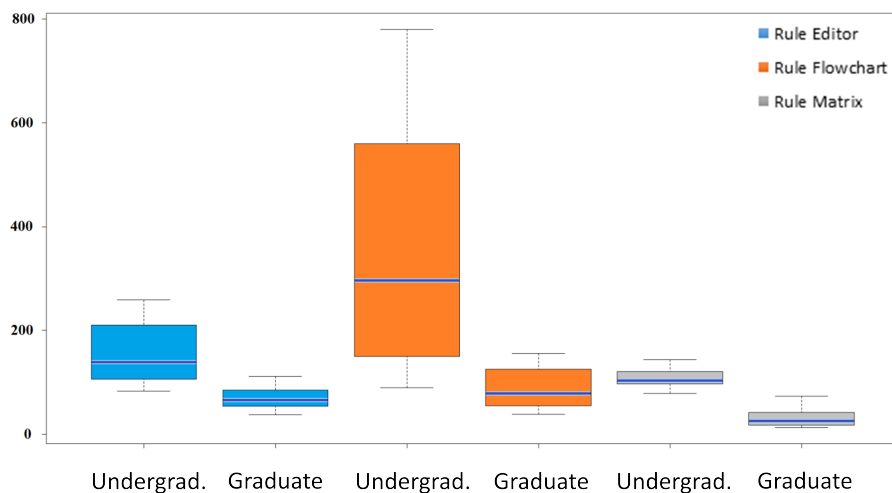


Figure 5.7: Duration (in s) of exercise 1 per education level (undergraduate or graduate).

We performed statistical tests on the results obtained with the different exercises, both in terms of time and number of errors. However, considering the population and the amplitude of the results, Wilcoxon tests did not yield significant results. We then applied ANOVA (analysis of variance). For this purpose, we conducted tests of homoscedasticity variance (Bartlett's test) indicating that the variances between our classes were significantly homogeneous. Finally, ANOVA results indicated a significant difference between the three tools with a confidence level of 10 %.

Figure 5.7 indicates the durations of the exercise 1 per level of education (undergraduate or graduate). The figure shows different statistical elements: the average is represented by a thick line contained in the box plot. The box extends from the first quartile to the third quartile and the whiskers extend to at least 1.5 times the inter-quartile range. The durations of exercise 1 were longer than the durations of the others because this exercise included the discovery of the tools by the participant who did not receive any outside assistance at this stage of the experimentation. These graphs show firstly the disparity

between undergraduate participants and graduate ones: the average duration is always longer. Then, the range of durations for undergraduate people is larger. In addition, the graphs show that flowchart is the most difficult tool to handle, especially for undergraduate people, while in contrast, everyone finished quickly the exercise with the association matrix. To understand these figures, we have to remember that there are a few ways to interact with the matrix: users just had to choose two inputs and select a value in each cell. It is therefore normal that the tool was mastered quickly. Finally, the durations with our tool were longer than for the association matrix, although average durations are comparable. Undergraduate participants had also found it more difficult to overcome *Rule editor* than *Rule matrix*.

The durations of the other exercises were quite similar except that the maximum durations were lower than those of exercise 1 (see table 5.1). The duration of exercise 3 with Rule matrix was just the time to understand it is not possible to use this GUI for the exercise. The association matrix remained the fastest tool to create rules, closely followed by our rule editor. The flowcharts were still slower, especially for undergraduate people (on average, there were more than 2 minutes between the two populations). In terms of population categories, experts in artificial intelligence had achieved the fastest all exercises regardless of the tool, followed by scientists and then by the other candidates. These results confirmed the intuition.

Exercise	GUI	Undergraduates	Graduates
Exercise 2	Rule Editor	0'58"	0'36"
	Rule Flowchart	1'34"	0'40"
	Rule matrix	2'44"	0'35"
Exercise 3	Rule Editor	3'09"	2'01"
	Rule Flowchart	4'52"	2'09"
	Rule matrix	0'48"	0'25"

Table 5.1: Average durations for exercise 2 and exercise 3 per education level.

In terms of speed of mastering and execution, nothing seemed faster than the association matrix. However, for now we have considered only the execution time of the exercises without looking at the errors produced by users. Statistical tests showed that the number of errors is independent of the education level.

Figure 5.8 shows the error rate of the whole panel on each exercise and for each tool. Errors were mostly logical misunderstandings (e.g. the use of a disjunction instead of a conjunction). Whereas association matrices were the fastest way to capture the rules, they got the highest error rate, which increased with the difficulty of the exercise. Browsing through input pairs to create the rules seemed to make the exercise difficult and required more concentration. We also counted as error the rules that were created when the user tried to write rules with more than two entries (which, we remind, is impossible). Finally, the rules were more difficult to read when displayed in a tabular form than in a textual form. This was also the case for flowcharts, which could also explain the error rate of the *Rule flowchart* tool. However, with the use of the *Rule editor* tool, the users made fewer mistakes: this can be explained by the fact that the rules are presented in an easily readable form, unlike the other tools. In the case of *Rule editor*, only one mistake was committed.

In a more subjective way, we asked the participants what was the tool they would prefer if they had to use daily. All undergraduate participants answered they preferred *Rule editor*. This made sense since our tool is not based on a mathematical or logical representation that these users were not used to. However, for graduate participants, the

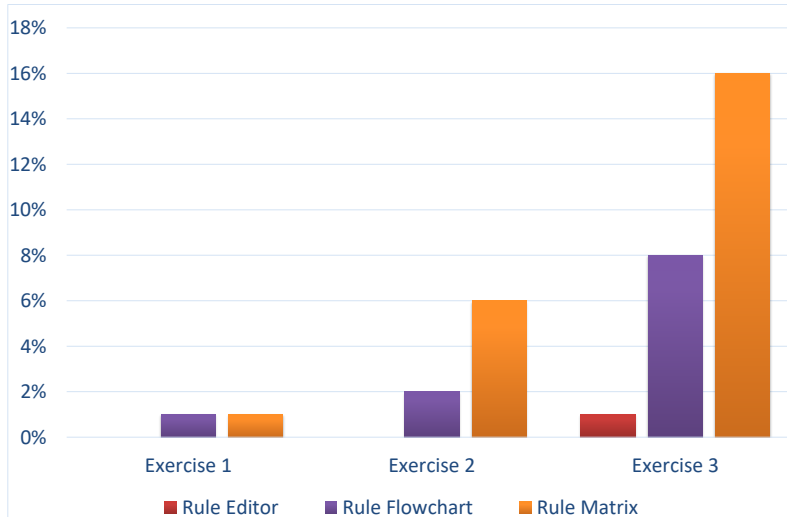


Figure 5.8: Error rate of all users, per exercise and per tool (in %).

answers were a bit more varied: 43 % responded *Rule editor*, 29 % *Rule flowchart*, 9 % *Rule matrix* and finally 17 % would use *Rule editor* for small and medium rules and *Rule flowchart* for building rules that are more complex .

We also compared the answers regarding the level of knowledge in fuzzy expert systems (expert, intermediate and novice): 60 % of the experts preferred *Rule editor*, while others were distributed equitably between the other two tools. No novice user has appreciated *Rule matrix*, despite what the quantitative analysis revealed; however 50 % of them have chosen *Rule editor* and 12 % a mix between *Rule editor* and *Rule flowchart*. Finally, 60 % of intermediate users have chosen *Rule editor*. All figures are reported in Table 5.2.

	Experts	Intermediate	Novices
<i>Rule editor</i>	60%	60%	50%
<i>Rule matrix</i>	20%	10%	0%
<i>Rule flowchart</i>	20%	10%	38%
<i>Rule editor+Rule flowchart</i>	0%	20%	12%

Table 5.2: User preferences for a tool per skills.

According to the experimental results, graduated people quickly and easily adopted association matrices. However, for people less familiar with mathematics and logic, it seemed to be less easy and the mistakes are numerous. In addition, the tool was very limited because only rules with two inputs and conjunctions could be edited.

Similarly, scientific people appreciated flowcharts, especially for complex rules. However, the reading was not easy and novice users were struggling to reason correctly and check their rules.

Finally, *Rule editor* seemed to suit the majority of users. The adoption of the tool and the rule authoring speed were similar to the association matrices. On the one hand, the textual form helped to achieve a low error rate. On the other hand, the time to author a rule was only slightly greater than with association matrices (if the rule could be supported by matrices). The GUI is still our main tool to edit rule bases. Over the years, it has changed graphically, but it is still relying on the same principles.

To improve the acceptance of fuzzy expert systems, the next section describes how we provide the users with an explanation of the decisions.

5.2 Generation of textual explanations of fuzzy systems decisions

The most important lever to increase the usability and the acceptance of fuzzy systems as XAI is the generation of explanations. Alonso *et al.* (Alonso Moral, Ramos-Soto, et al., 2017) emphasize that users prefer decisions accompanied by an explanation. Such explanations are now called local explanations in opposition to global explanations that are related to the model itself.

Very soon after they appeared, expert systems offered textual (local) explanations: they relied on templates that were completed with some elements from the trace of the reasoning (Moore and Swartout, 1989). The advances in Natural Language Generation allow coupling systems together to benefit from a more powerful framework to create those explanations. In our work, we use *SimpleNLG*, originally developed by Ehud Reiter (Gatt and Reiter, 2009). It is a free API to function as a realization engine, i.e. a program that creates a well-formed sentence from its constituents.

In our work, we focused on two case studies. The first one involves a fuzzy classifier. The goal of this work was to develop a complete, yet simple, workflow for textual explanations. We designed an approach from the extraction of the main elements from the trace of the reasoning to the evaluation of the explanations by a human panel. The second case study is about the generation of explanations of semantic annotation with the model trained as described in section 4.3. The latter approach relies on expressiveness as defined in chapter 3 and on insights from cognitive science and psychology.

5.2.1 First case study: fuzzy classifier

The first case study was the topic of Ismaïl Baaj's internship, just before his PhD. The goal of the internship was to set up a complete workflow, from the trace to the textual explanation, and to consider the evaluation of this explanation. The underlying purpose was to identify concretely and empirically the scientific locks. We limited ourselves to the simplest fuzzy system: a fuzzy rule-based classifier (FRBC) as presented in chapter 2, section 2.5.2. FRBCs are a good first case study because they do not use aggregation and defuzzification methods that complicate the interpretability, and thus affect the explainability.

We designed a method consisting of three steps as shown in figure 5.9. The first step consists in selecting from the execution trace T the fuzzy propositions that are the most relevant for the explanation. Let us define $T = \langle KB, M \rangle$ the execution trace of a fuzzy classifier where:

- KB is the knowledge base, i.e. a minima the rule base.
- M is a mapping between any object (any expression, any rule) and its value.

Note that KB is global, in the sense it is the same for any instance to be classified, whereas M is local and depends on the current instance.

The second step consists in decorating the fuzzy propositions with language moderators (e.g. "more or less", "extremely", etc.) to get the sentences constituents. Finally, the generation of text is mainly realization and gives as output the textual explanation.

Justification extraction

In this first step, we receive as input the execution trace of the inference engine. In ExpressIF[®], it is possible to query the trace T for the value of any expression, input, rule etc. The goal is to select the parts of the premises that will be useful in the explanation: this is why we called that "reduction". In the end, we have a set of fuzzy expressions that

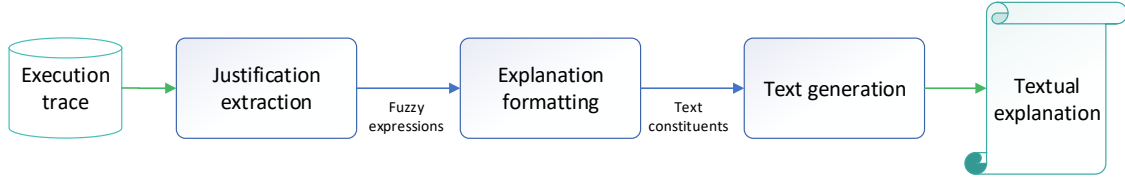


Figure 5.9: Overview of the proposition of textual explanation generation.

is close to a conjunctive form (like the Conjunctive Normal Form in Boolean logic), except that the conjunction is not the logical one but its linguistic counterpart.

Definition 5.1 (Reduction function \mathcal{R})

Let e, e_1, e_2 be fuzzy expressions, $\alpha_e, \alpha_{e_1}, \alpha_{e_2}$ their respective fuzzy values. Let \mathcal{R} be the reduction function that takes a fuzzy expression as input and outputs fuzzy expressions. We define \mathcal{R} recursively:

1. If e is a fuzzy proposition

$$\mathcal{R}(e) = e. \tag{5.1}$$

2. If e is a conjunction of the form $e_1 \wedge e_2$:

$$\mathcal{R}(e) = \begin{cases} \mathcal{R}(e_1) & \text{if } \alpha_{e_1} = 0 \text{ and } \alpha_{e_2} \neq 0 \\ \mathcal{R}(e_2) & \text{if } \alpha_{e_1} \neq 0 \text{ and } \alpha_{e_2} = 0 \\ (\mathcal{R}(e_1), \mathcal{R}(e_2)) & \text{otherwise.} \end{cases} \tag{5.2}$$

3. If e is a disjunction of the form $e_1 \vee e_2$, let $Th \in [0, 1]$ be a threshold that can be arbitrarily $Th = 0.75$, and

$$\mathcal{R}(e) = \begin{cases} (\mathcal{R}(e_1), \mathcal{R}(e_2)) & \text{if } \alpha_{e_1} = 0 \text{ and } \alpha_{e_2} = 0 \\ \mathcal{R}(e_1) & \text{if } \alpha_{e_1} \neq 0 \text{ and } \alpha_{e_2} = 0 \\ \mathcal{R}(e_2) & \text{if } \alpha_{e_1} = 0 \text{ and } \alpha_{e_2} \neq 0 \\ \mathcal{R}(e_1) & \text{if } |\alpha_{e_1} - \alpha_{e_2}| \geq Th \text{ and } \alpha_{e_1} > \alpha_{e_2} \\ \mathcal{R}(e_2) & \text{if } |\alpha_{e_1} - \alpha_{e_2}| \geq Th \text{ and } \alpha_{e_2} > \alpha_{e_1} \\ \mathcal{R}(e_1) \vee \mathcal{R}(e_2) & \text{otherwise.} \end{cases} \tag{5.3}$$

Intuitively, in the second case (conjunctions), we know that if only one operand equals 0, it is responsible for the value of the conjunction. Otherwise, we cannot conclude anything so we just try to reduce both operands but we keep the conjunction as part of the explanations.

The third case (disjunctions) can be interpreted as follows. If $\alpha_{e_1} = \alpha_{e_2} = 0$, both expressions participate to the nullity of α_e : we thus transform e into a linguistic conjunction in the explanation. Then, if only one of the two operands equals 0, only the other explains α_e . The threshold is used to simplify the disjunctions in the explanation: the intuition behind is that if one of the two operands is really preponderant in the value α_e , the other operand is not necessary. In all other cases, we keep the disjunction, simplifying its operands.

Note that we did not talk about negations because they are rare in such systems. Anyway, if a negation should apply to a fuzzy proposition, it is equivalent to taking the complement to 1 of the membership function of the term. If the negation applies to a conjunction or a disjunction, then we consider with De Morgan triplets (see chapter 2 for

a definition) that it can be simplified and we go back to the previous case. Finally, we consider the other cases are degenerate.

To avoid confusion with the explanation itself, we introduced the notion of justification. A justification is rougher and is a set of fuzzy expressions, eventually ordered. The justification of a rule conclusion c is denoted as $Justification(c)$, and the justification of the whole decision is simply denoted as $Justification(\mathcal{D})$. We distinguished three cases of classification with the fuzzy classifier that contains N rules:

- Simplest case: only one rule r^* is more activated than the others and its activation is high. The classifier performed well, the explanation is straightforward. Let e^* and c^* be respectively the premise and the conclusion of p^* , then:

$$Justification(\mathcal{D}) = Justification(c^*) = \mathcal{R}(e^*). \quad (5.4)$$

- Ambiguous case: several rules have the same high activation (to an epsilon). The explanation must be transparent about this hesitation. Let us consider the ordered set of such rules $\{r_i\}_{i \in [1;M]}$ with $M \ll N$ such that

$$\forall j, k \in [1; m], j \neq k, |\alpha_{r_j} - \alpha_{r_k}| \leq \epsilon \quad (5.5)$$

and such that if $j < k$ then $\alpha_{r_j} \geq \alpha_{r_k}$. ϵ is a constant such that $0 \leq \epsilon < 1$ that must be close to 0. In that case, each different class will have its justification, and we keep the order regarding the rule activation. Let e_i and c_i be respectively the premise and the conclusion of the rule r_i .

$$Justification(c) = \bigcup_{c_i=c} \mathcal{R}(e_i)$$

Let $\mathcal{C}' \subset \mathcal{C}$ the set of classes in $\{r_i\}_{i \in [1;M]}$:

$$Justification(\mathcal{D}) = \{Justification(c), \forall c \in \mathcal{C}'\}. \quad (5.6)$$

- Worst case: only one rule r^* , whose premise and conclusion are respectively denoted by e^* and c^* is activated with a low activation. The classifier does not perform well, and we have to tell the user that the decision is not sure at all, while helping him. In the terminology of Stepin *et al.* (Stepin et al., 2021), we need a contrastive-counterfactual explanation. Let r denote a rule and e and c respectively its premise and conclusion (i.e. class). As before, we have:

$$Justification(c) = \mathcal{R}(e).$$

Let us now define the set of elements $Counterfactual(\mathcal{D})$ that will constitute the contrastive-counterfactual part of the explanation. Let us consider the set of unactivated rules $\{r_i\}_{i \in [1;M]}$ with $M < N$ such that $\forall i \in [1; M], \alpha_{r_i} = 0$ and $\mathcal{C}' \subset \mathcal{C}$ the set of classes in $\{r_i\}_{i \in [1;M]}$:

$$Counterfactual(\mathcal{D}) = \{Justification(c), \forall c \in \mathcal{C}'\}. \quad (5.7)$$

Then,

$$Justification(\mathcal{D}) = Justification(c^*) \cup Counterfactual(\mathcal{D}). \quad (5.8)$$

Explanation formatting

Explanation formatting consists in minimizing and factorizing $Justification(\mathcal{D})$. The justification of a class (conclusion) $c \in \mathcal{C}$, denoted as $Justification(c)$, is going to be realized as a conjunction of phrases, notably with coordinating conjunction "and". It is important to notice that the justifications do not contain any negation of binary expressions, with the help of \mathcal{R} . During the realization, we will enunciate conclusions from the most possible to the least to preserve consistency. Thus, the sort is performed regarding the activation degrees in descending order.

Text generation

The goal is to enrich the explanation with useful characteristics of the situation, by finding for each conclusion c an accurate linguistic terminology regarding the fuzzy proposition values. For each fuzzy proposition, a qualifier regarding its value is added. We adapted the uncertainty scale described by Budescu *et al.* (Budescu, Por, and Broomell, 2012), presented in table 5.3, to consider one more situation: absolute certainty. Moreover, when a fuzzy proposition is evaluated to 0, we add the negation "not".

Phrase	Likelihood of occurrence/outcome
Certainly	>99%
Very likely	>90%
Likely	>66%
About as likely as not	33% to 66%
Unlikely	<33%
Very unlikely	<10%
Exceptionally unlikely	<1%

Table 5.3: IPCC uncertainty scale (Budescu, Por, and Broomell, 2012).

The items in *Counterfactual*(\mathcal{D}) are treated a little bit differently and a sentence is built to claim the impossibility of the given classes.

Result

Let us consider a FRBC that can classify pasta. The choice of the application domain is due to the human-based evaluation: it must be simple enough so that all respondents can feel familiar to the problem.

In our problem of pasta classification, we consider 5 input linguistic variables:

- Length is continuous with 3 terms (very short, short and long).
- Width is continuous with 5 terms (hair thin, very thin, thin, medium and large).
- Longitudinal profile is categorical with two terms (straight and twisted).
- Cross-section is categorical with two terms (hollow and solid).
- Surface aspect is categorical with two terms (smooth and striated).

In the set \mathcal{C} , we consider 8 types of pasta: Bucatini, Capellini, Fusilli, Linguine, Maccheroni, Penne, Spaghetti and Ziti. The rule base has been created mainly using a specialized book (Legendre, 2011). The terms of the continuous variables have been designed empirically by ... asking all the colleagues to measure pasta at home! The pasta rule base is presented in table 5.4. Each line of the table is a classification rule. For instance, the first row can be interpreted as:

IF length IS long AND width IS thin AND longitudinal profile IS straight AND cross-section IS hollow AND surface aspect IS smooth THEN class IS Bucatini.

In this example, the vocabulary \mathcal{V} of the fuzzy system is limited to the linguistic variables themselves (no specific predicates are used). In this context, we show hereafter an example of explanation for each cases we identified before:

- a very sure result:

Length	Width	Longitudinal profile	Cross-section	Surface aspect	Class
Long	Thin	Straight	Hollow	Smooth	Bucatini
Long	Hair thin	Straight	Solid	Smooth	Capellini
Very short	Large	Twisted	Solid	Smooth	Fusilli
Long	Thin	Straight	Solid	Smooth	Linguine
Short	Medium	Straight	Hollow	Striated	Maccheroni
Short or very short	Medium	Sheared	Hollow	Striated	Penne
Long	Very thin	Straight	Solid	Smooth	Spaghetti
Long or short	Medium	Straight	Hollow	Smooth	Ziti

Table 5.4: The pasta classifier rulebase.

This is definitely a Fusilli because longitudinal profile is twisted and cross section is solid and surface is smooth and length is definitely very short and width is definitely large.

- an ambiguous situation with two conclusions being possible:

This is likely a Capellini because longitudinal profile is straight and cross section is solid and surface is smooth and length is likely long and width is likely hair thin. There is another choice: it can be unlikely a Spaghetti because longitudinal profile is straight and cross section is solid and surface is smooth and length is likely long and width is unlikely very thin.

- one very unlikely result and all the reasons why each other conclusions are impossible:

This is exceptionally unlikely a Spaghetti because longitudinal profile is straight and cross section is solid and surface is smooth and length is exceptionally unlikely long and width is definitely very thin. Some conclusions are not possible:

- *It is impossible to be a Linguine because width is not thin.*
- *It is impossible to be a Capellini because width is not hair thin.*
- *It is impossible to be a Bucatini because cross section is not hollow and width is not thin.*
- *It is impossible to be a Maccheroni because cross section is not hollow and surface is not striated and length is not short and width is not medium.*
- *It is impossible to be a Fusilli because longitudinal profile is not twisted and length is not very short and width is not large.*
- *It is impossible to be a Penne because longitudinal profile is not sheared and cross section is not hollow and surface is not striated and length is not short and length is not very short and width is not medium.*
- *It is impossible to be a Ziti because cross section is not hollow and width is not medium.*

We will see in section 5.2.3 how a panel perceives these kinds of explanations.

5.2.2 Second case study: semantic annotation

The second case study is related to Regis Pierrard’s PhD and concerns another type of reasoning: Fuzzy Constraint Satisfaction Problems (FCSP) solving. We limited the problem to semantic annotation. Indeed, the algorithms presented in section 4.3 are used to generate a FCSP and are applied to MRI images annotation. In this work, we relied on the linguistic aspect of the spatial relations, discussed in chapter 3. We also limited to FCSPs that have at least one solution. The explanations for CSPs or FCSPs that have no solutions is a specific problem that is considered as difficult (Dev Gupta, Genc, and O’Sullivan, 2021).

A base algorithm for textual explanation generation can describe region by region (i.e. organ by organ) the mutual relations between them. Such an algorithm is described in Regis’ manuscript (Pierrard, 2020). We went further in a work with Wassila Ouerdane (MICS, CentraleSupélec), considering insights from other research fields.

Cognitive science considerations

Cognitive science has largely studied how Humans represent a scene or scan images. Thus, it seems natural to consider those insights to create an explanation.

Zwaan *et al.* present more than a decade of studies about situation model, i.e. a mental representation of affairs (Zwaan and Radvansky, 1998). They highlight the difficulty to describe correctly a spatial scene with language, because of the difference between its dimensionality and the dimensionality of space. For instance, if one describes a room in a circular way, the first and the last objects are far from each other in the description but close together in the room. This also shows the importance of the order in which the parts of the scene have to be described.

This leads us to the studies about image scanning (Borst, Kosslyn, and Denis, 2006), which is related to the mental representation of a scene or an image. In (Kosslyn, Ball, and Reiser, 1978), authors state that the visual images preserve the metric spatial information. Indeed, subjects need more and more time to visualize mentally the information when going further to the focus point. Other works study the difficulties of subjects to represent a scene if the description is too long and if the description is too precise (Denis, Goncalves, and Memmi, 1995; Farah and Kosslyn, 1981). Another difficulty is the direction of reading: (Román, Fathi, and Santiago, 2013) indicates that it affects the description of a scene.

The studies about image scan paths bring also good information. The attention of subjects is classically attracted by focus points. In image understanding, this is called salient objects and (Borji *et al.*, 2014) gives a comprehensive review on their automatic detection. Nevertheless, cognitive science warns of the difficulty of defining saliency because it can be context-dependent, or due to the singularity of an object, of the user’s goal, etc. However, when the same subject watches the same picture, the scan paths may be different (Noton and Stark, 1971): thus, the scan path does not depend only on the objects in the image. If several similar pictures are presented, the scan path can also be more and more efficient (Noton and Stark, 1971).

Finally, the Gestalt psychologists (Luccio, 1993) studied the cognitive issues of visual perception, in particular the shape of objects. The 7 Gestalt principles concern figure-ground, similarity, proximity, common region, continuity, closure and focal point of images. They are particularly useful in design, but give some insights about how objects are perceived. In particular, they recommend grouping objects that are similar or that share properties.

This short overview of cognitive science helped us to design our explanation strategy, whose main principles are:

- *Sorting*: the order of the results has an importance. It is important to start with regions in images that are salient, and then, regarding the recommendations of cogni-

tive science papers, use diagonals and increasing distances to select the next results. The spiral order is not recommended.

- *Saliency*: the saliency is a difficult concept that can be context-dependent. *A minima*, one can select the biggest object or a group of objects as focus point.
- *Symmetry*: a pair of objects that are symmetrical must be grouped.
- *Priority*: we must select the most satisfied constraints first.
- *Associativity*: some relations are associative (e.g. "to the left of") and explainees can immediately infer it, so we must use that to reduce the number of constraints involved in the explanations.
- *Locality*: if possible, we will use first the constraints with the closest regions in the image.

Moreover, an explanation must somehow indicate how the task has been achieved. In our case, the solving of a FCSP is quite simple to explain since the algorithm searches for the values of the variables such as the constraints are satisfied. However, it makes the explanation more complicated when constraints are not all unary, since these assignments are dependent from each other. Indeed, for instance, a binary constraint will force the assignment of two variables together. In the case of semantic annotation or classification, the constraints are relations so that it is a little bit simpler than, for instance, quadratic constraints.

Another point is that we need to select a maximum number of constraints for each variable. In addition, there must be no correlation between these constraints: for instance, the values of "to the left of" and "to the very left of" may be correlated and so we do not want to use them at the same time for the same variable because they are redundant. We use mutual information to detect this correlation.

Description of the algorithm

Let us now present the algorithm to generate concise explanations for semantic annotation from the execution trace of the solver. Let $T = \langle P, s, \bar{C}_s \rangle$ be this execution trace that is composed of:

- $P = \langle X, D, C \rangle$ is a FCSP.
- s , a chosen solution among all the solutions of P , for instance, the best one regarding the degree of consistency. s contains the assignment for each variable in X .
- \bar{C}_s , the set of degrees of satisfaction of each $c \in C$ in the solution s .

The explanation starts with a general sentence that indicates the global confidence about the annotation based on the degree of consistency of the solution. Then, we select the region from the segmentation that is the most salient. Regarding this object, the image is divided into four quadrants. The explanation will start with the most salient region, then with the other objects in the same quadrant, then quadrant by quadrant, in the clockwise order.

For each variable of the FCSP, we have to select at most N_{max} constraints to justify the explanation. The constraints are chosen regarding not only their level of satisfaction (that must be the highest as possible not to overload the text with moderators), but also their mutual link and the proximity with the other variables. The mutual link between relations is a tricky part. We use a knowledge graph about the relations as presented in chapter 3, section 3.5. Such a graph emphasizes different links between two relations r_1

and r_2 . In particular, the symmetry link is important not to use twice the same constraint. Let o_1 and o_2 be two objects in the image, and r a symmetrical relation, if $o_1 r o_2$ is used in a sentence, we cannot use $o_2 r o_1$ anymore.

Then, special attention is given to grouping constraints such as "is symmetrical to" that constitutes a pair of variables. Indeed, the previous section highlights that groups of objects must be processed together. Thus, the other variables in the scope of this constraint must be processed just after.

Results

In this example, the FCSP has been extracted automatically from few images from the Visceral dataset¹ with the approach described in section 4.3. Figure 5.10 shows one of the images and different organs of interest.

The segmentation has been obtained automatically and the regions were given an identifier in an arbitrary order. However, for the sake of comprehension of this example, we numbered the organs, from left to right and top to bottom.

The vocabulary \mathcal{V} is limited to 9 spatial predicates (as seen in chapter 4, section 4.3.6):

*ToTheLeftOf, ToTheRightOf, Above, Below, IsSymmetricalTo,
CompletelyToTheLeftOf, CompletelyToTheRightOf, CompletelyAbove,
CompletelyBelow.*

Let consider the solution of such a FCSP applied to figure 5.10 with the highest degree of consistency.

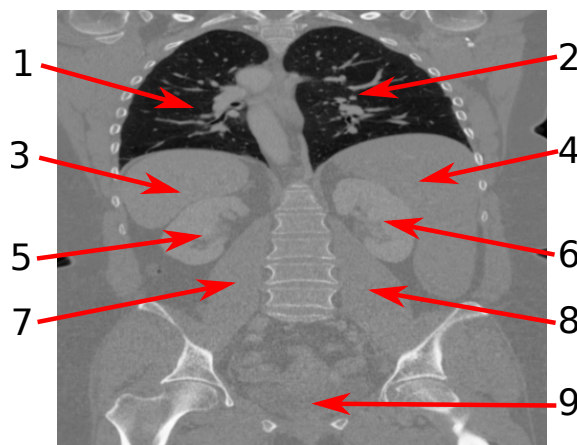


Figure 5.10: Backward MRI image with different regions to annotate.

For the sake of simplicity, we defined the focus point as the biggest object (in terms of area). We set the constant $N_{max} = 2$. Most of the constraints are linked in the knowledge graph, because we used mainly directional relations like "at the right of" and "at the left of". This explains why we rarely reach N_{max} constraints.

The algorithm presented just before generates the following explanation:

¹<http://www.visceral.eu/>

*This is the annotation of the given image (with a very high confidence).
The right lung (region 2) is symmetrical to the left lung (region 1) and
above the liver (region 4).
The liver (region 4) is to the right of the right kidney (region 6) and to
the right of the right psoas (region 8).
The right psoas (region 8) is above the bladder (region 9) and is sym-
metrical to the left psoas (region 7).
The left psoas (region 7) is below the left kidney (region 5).
The spleen (region 3) is above the left kidney (region 5) and is below the
left lung (region 1).*

It is important to note that this algorithm is not domain-specific. Indeed, the relations are generic in the sense that they could be used in another domain (such as satellite image annotation). It also manipulates image regions and has no clue they represent organs. However, the labels that are used are organ names, because we want a semantic annotation. We do not use external domain knowledge, for instance, to replace the word "region" with "organ" in the explanation, or to use a more technical vocabulary.

The next section presents an evaluation of this kind of explanation.

5.2.3 Human-based evaluation of explanations

The last contribution in the field of textual explanation concerns the evaluation. Today, two main families co-exist: objective and human-centered evaluation approaches (Vilone and Longo, 2021). For the latter one, the evaluation requires the definition of a test protocol and of criteria that characterize a good explanatory capacity and therefore the quality of the explanations (Doshi-Velez and Kim, 2017).

From the literature, we can distinguish three main topics to characterize an explanation:

- Natural language: its evaluation resides nowadays in the evaluation of a NLG layer as suggested by Alonso (Alonso Moral, Ramos-Soto, et al., 2017). Reiter and Belz described ways to assess the quality of text produced by NLG systems (Reiter and Belz, 2009), but in the case of explanations, many choices like the tense need also to be studied.
- Human-Computer Interaction: after reading these explanations, the interaction between the human and the intelligent system and their relation may evolve. For instance, the user can change his own opinion if he or she is convinced by the system.
- Content and form need strong features from the way humans use explanations to communicate. From a social sciences point of view, Miller argues that the most important criteria are probability, simplicity, generality and coherence with prior beliefs (Miller, 2019). Except for the coherence that has been defined by Thagard (Thagard, 1989), criteria do not have a well-developed formalism.

We proposed a survey questionnaire with 17 assertions shown in table 5.5 and assessed with a Likert scale (from 1 "strongly disagree" to 5 "strongly agree"). We added a comment part for each, as suggested by Moore who declared that they are often the most interesting parts to understand the frustrations of the user with the system, and help to improve explanations (Moore, 1994).

We use this questionnaire on the two case studies we presented before.

Evaluation of explanations in the first case study (fuzzy classifier)

The panel was composed of 69 respondents, 9 of them working in the field of Artificial Intelligence. The questionnaire has been broadcast on the web and displayed the pictures

Natural language	<ol style="list-style-type: none"> 1. Overall, explanations are written in a correct English 2. Conjugation choices are appropriate and adequate 3. Grammatical form of sentences is satisfying
Human-Computer Interaction	<ol style="list-style-type: none"> 4. Explanations are simple to use and easy to read 5. Explanations help to make decisions faster than without 6. Explanations let you change your opinion about your expectations 7. Explanations help to make good decisions and are convincing 8. Data and explanations are enough to trust the system 9. Explanations express indirectly the way of the system is reasoning
Content and form	<ol style="list-style-type: none"> 10. Length of explanations is adequate 11. Explanations are not repetitive 12. It is easy to read explanations until the end 13. Content layout and order of elements in explanations are satisfying 14. All causes are identified in explanations 15. Explanations do not contain any superfluous information 16. Overall, explanations seem consistent 17. Explanations are true

Table 5.5: Survey questionnaire to evaluate textual explanations.

of each of the 8 different pasta before any assessment. It contained the three explanations presented before.

Figure 5.11 presents the results of the evaluation. The red color indicates disagreement, in opposition to the blue color that indicates agreement. In one look, it is possible to see that the blue color is dominating, i.e. the evaluation is globally positive, except for the repetitiveness (mainly due to the third kind of explanation).

In terms of natural language, users were satisfied with the English writing (74% approved) and the choice to use the present as conjugation (78% agreed). However, they were more undecided about the grammar as 39% disapproved and 48% approved that it was correct enough. They notably said that the explanations were unusual (too much *and*) but remained understandable. A lack of punctuation has been pointed out, and some morphology mistakes have been noticed (e.g. *can be unlikely* was not correct...). In addition, few people find unnatural the qualifiers for uncertainty and the variable named "longitudinal profile". The fact that the algorithm adds systematically a qualifier made the sentences heavier. This point must be improved in a future version.

We also noticed that some users encountered difficulties while reading the explanations (as 39% people disagreed and 48% others agreed with assertion 4 and only 49% agreed with assertion 5) but all agreed that they strongly expressed the way our system reasons (78% people agreed with assertion 9). Users are sufficiently convinced by explanations (only 18% surveyed disagreed with assertion 7) but felt there were not enough cases to fully trust the system (only 49% people agreed with assertion 8), i.e. the test seemed too short to them. Following the comments, assertion 8 and especially assertion 6 were sometimes misunderstood and subject to personal interpretation of the question. This was a difficulty when using a survey questionnaire: we did not want to take too much time to the panel. There was also a balance to find here between the time a user could spend and the willingness to ask many questions. For instance, in the beginning, we wanted to

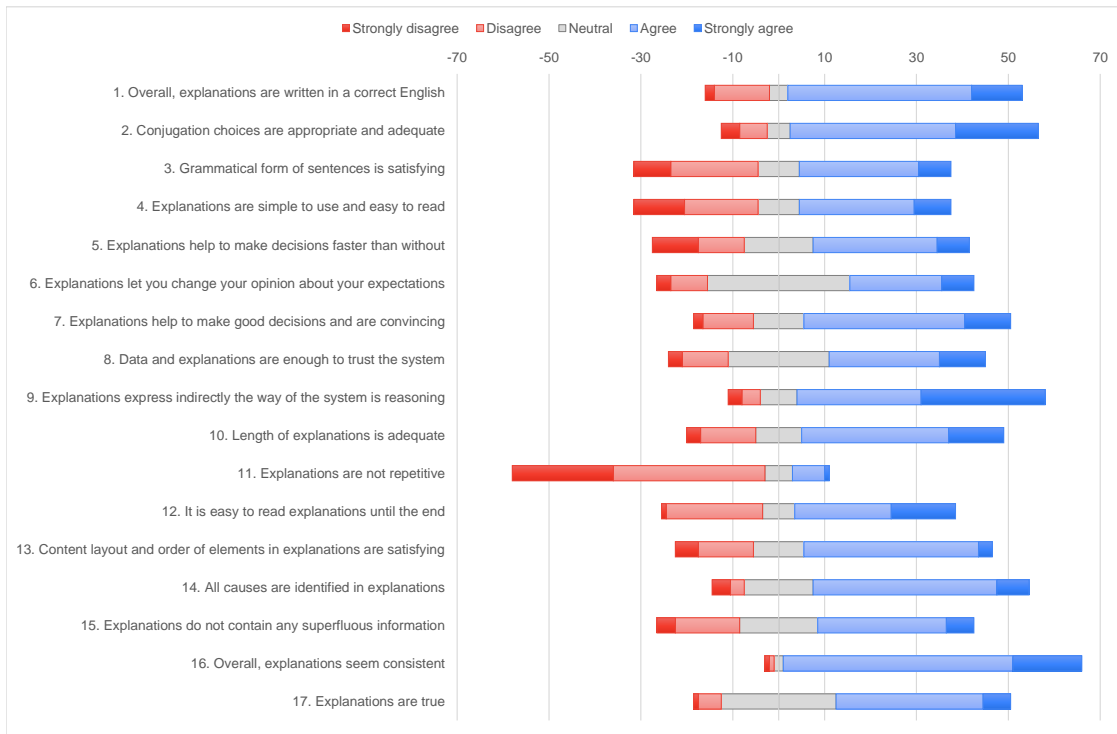


Figure 5.11: Results of the evaluation of the explanations generated in the pasta case study.

question about the 17 assertions for each explanation.

Even if the length of the explanations seemed satisfying to the users (64% agreed with assertion 10), they argued that they were too repetitive (80% of them for assertion 11). Nonetheless, the provided explanations were perceived as extremely consistent (94% agreed with assertion 16), and all causes were correctly identified to most of the people (68% agreed with the 14th fact, only 10% disagreed).

Evaluation of explanations in the second case study (semantic image annotation)

In this case study, the panel consisted of 40 respondents, with 20 medical staff members (medical doctors, surgeons, nurses, radiologists), the other half being computer scientists (6) and other various non-medical professionals (14). To decrease the medical staff's amount of time dedicated to the questionnaire, we selected only 12 assertions out of the 17 initial ones. In particular, we removed the assertions about the grammar (assertions 2 and 3), and other assertions that were not relevant to the use case (and the panel): assertions 5, 6 and 14. Figure 5.12 shows the results of the evaluation by the panel of the explanation presented before. Again, the red color is used for the disagreement and blue for the agreement. Globally, the evaluation is positive (even more positive than the previous case study). We did a statistical study to find out if the answers of the medical staff members were different from the others. It turned out that there was no significant difference between the responses of the two subgroups, that is why we considered the panel as a whole.

Seemingly, we improved our skills in using SimpleNLG since 95% of the respondents thought the syntax was correct. The comprehension of the reasoning behind the annotation (60% agree) and the transcription of the uncertainties (65% agree) were satisfying. These assertions showed that the respondents did not perfectly understand how the algorithm

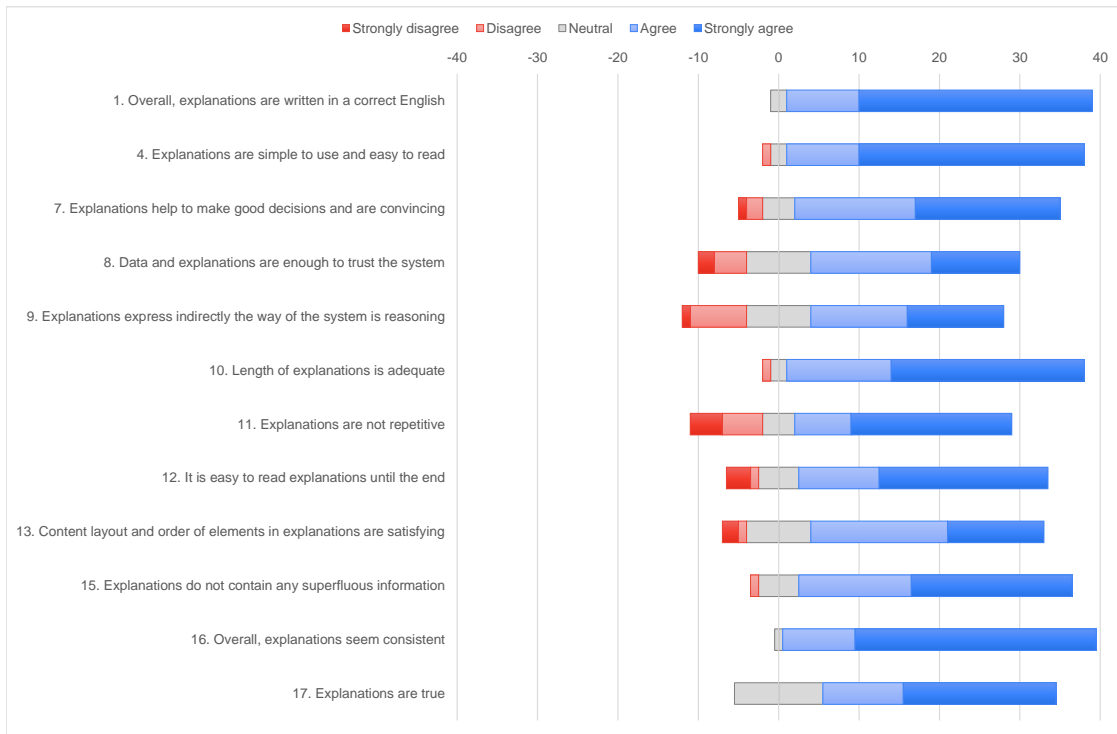


Figure 5.12: Results of the evaluation of the explanation generated in the MRI semantic annotation case study.

annotated the organs and why the algorithm was not confident in all the cases. For the other assertions, the results were very good: 92% of the respondents thought the explanation was easy to read and that its length was adequate (assertions 4 and 10). The answers were more nuanced in terms of repetitiveness (assertion 11) since 22% of the panel thought the explanation was repetitive. Finally, 65% of the respondents thought they could trust the annotation.

The questionnaire invited also the respondents to write comments after each type of explanation. Most of the medical staff felt uncomfortable with the fact that the MRI image was taken from the back. Nevertheless, no one declared the explanation was wrong: maybe it could have an impact on the confidence of the users in the AI.

One of the medical respondents said it could be useful to use the spine as the main region and use it for the labeling of the other regions. This idea emphasized the importance of saliency: indeed, in such an image, we can see the spine first because it is whiter and central. Unfortunately, in the segmentation we used, bones were not considered.

We have also compared this kind of explanation and the raw explanation described in (Pierrard, 2020): obviously, taking into account the way humans perceive the explanations led to better explanations (Poli, Ouerdane, and Pierrard, 2021a).

5.3 Summary

In this chapter, we tackled the problem of the usability of fuzzy systems to bring them closer to XAI regarding two axes.

The first axis concerns the way an end-user can edit the rule base, or more generally the knowledge base. Rule base edition is useful even if the rule base has been induced by machine learning approaches, e.g. to adapt some parameters or if an expert wants to correct the rules. We studied the cognitive impact of such an activity and we proposed a graphical user interface that allows editing both vocabulary and rules seamlessly. We

compared this new tool with classical tools to edit rules with a Human-based evaluation and showed our *Rule editor* is preferred to the others.

The second axis is about the generation of textual explanations. We developed algorithms for two case studies concerning classification and semantic annotation. In both cases, we relied on Natural Language Generation, specifically the *SimpleNLG* realization engine. On the one hand, the classification is performed by a fuzzy classifier. We generated different kinds of explanations regarding the number of activated rules. We introduced a "reduction" function, which reduces the trace of the reasoning to the minimal set of items that explains the decision. On the other hand, the semantic annotation relied on a fuzzy constraint satisfaction problem solving. To improve the quality of the explanation, we used insights from cognitive science and psychology, decreasing the repetitiveness and complexity of the first explanations.

We also developed a method to evaluate such explanations that belongs to the Human-based approaches. It is based on the choice of a relevant panel and a questionnaire composed of 17 assertions. This methodology has been applied to both case studies and an analysis of the results has been proposed.

So far, we have presented the three main kinds of contributions to elevate fuzzy systems to the rank of XAI. We also tackled a last problem, trying to reduce the distance between fuzzy systems and the sensors that produce data. This is developed in the next chapter.

Chapter 6

Bringing fuzzy systems closer to digital instrumentation

6.1	Fuzzy logic for digital instrumentation	124
6.2	Using knowledge to build interpretable features	125
6.2.1	SAW sensors and diamond functionalization	125
6.2.2	Estimation of the contributions parameters as new features	127
6.2.3	Towards the recognition of mixture of compounds	128
6.2.4	Application	128
6.3	Adapting fuzzy decision trees to specific data	130
6.3.1	Context	130
6.3.2	Learning vocabulary from uncertain data	131
6.3.3	Fuzzy decision tree induction from uncertain data	132
6.3.4	Decision making	133
6.3.5	Empirical comparison	134
6.3.6	Application	134
6.4	Summary	138

In this last chapter of this document, we present the last kind of contributions. CEA is a research institute where you can meet engineers and researchers from different disciplines. I am always open to apply our research on their fields. In particular, for more than one decade, I have worked with physicists and chemists on digital instrumentation. This allows being closer to the data source, with all the drawbacks that we can imagine about the noisy aspect of real world measurements. Nevertheless, XAI should be applied to this kind of data while being still offering interpretability and explainability. This is particularly true because most of the applications we are interested in regarding digital instrumentation (namely security, health, manufacturing, etc.) are human-centered and the final decision will be made by an end-user who must be convinced, without being an AI professional.

The contributions in this field are divided into two parts. On the one hand, the first part may seem thankless and consists in developing softwares to manipulate these new sensors. In particular, this also gives access to their data. The other part consists in analyzing these data to make decisions. The purely technical part forms what is called ExpressIF Sensors[®], while the other part allows upgrading ExpressIF[®]. These contributions are motivated by my earlier experience with instrumentation. Broadly, when it comes to classifying samples, physicists and chemists use an approach that is similar to k-Nearest Neighbors (mainly with $k = 1$). They acquire few measurements, in general several per class to have some

variability in their data. They then use a distance to classify the current instance. We can also find neural networks, even if the number of training examples is often not sufficient. We started Olivier’s PhD with a huge benchmark of the methods classically used in the field of chemical sensors (Hotel, 2017; Hotel, Poli, Mer-Calfati, et al., 2018). More recently, with the availability of well-packaged toolboxes, XGBoost and Random Forests are also used. It should be possible to use fuzzy rule bases to enhance the performance of such approaches.

In this chapter, I will focus on two main scientific contributions. The first one is mainly the PhD of Olivier Hotel. The use of physical knowledge led to better features that are interpretable and more discriminative than the classical features extracted from signals. Chronologically, this work motivated the work on automatic feature construction presented in chapter 4, section 4.2. Then, the second contribution is about modifying the existing algorithms for fuzzy decision trees to handle specific data from instrumentation.

I will start with a short plea on the interest of fuzzy logic for digital instrumentation.

6.1 Fuzzy logic for digital instrumentation

Digital instrumentation is the field of digital sensor conception. They are step by step replacing their analog counterparts. Digital means that at one step, the analog phenomenon is changed into a digital output. The main principle of ExpressIF Sensors[®] is to be able to get four kinds of data from sensors and systems:

- Signals are related to event and data streams, whose distinction has been discussed in chapter 3. They are obtained when the sensors acquire measurements over time.
- Spectra are measurements along one or more continuous variables (e.g. the energy in nuclear science).
- Images can be viewed as arrays of intensity along one or more channels.
- Scalars are the simplest case.

Any of these types of data can be split into different kinds of granules (see chapter 2 for a definition). For instance, images can be split into regions, edges, spectra and signals can be split into peaks, which can have attributes like their position, their width, their height. Granulation concerns also the domains of scalars that can be split into terms of a linguistic variable. The granulation is thus an important aspect in the processing of sensors data. The organization can also be an important aspect in digital instrumentation data analysis: for instance, the position of the peaks in energy spectra are useful for isotopic identification or the relative amplitudes of multi-valued signals. Therefore, we retrieve here two aspects of human intelligence as viewed by Zadeh (Zadeh, 1997) and discussed in chapter 2.

The granulation of such data is carried out using signal and image processing methods, of which the literature abounds. In the particular case of spectra, it is necessary to deal with the calibration step that depends on the nature of the spectra. Broadly, the calibration is the transformation of the X-axis from channels into a discretized continuous variable as the energy.

Any system or sensor makes errors in the measurement of the physical quantities or is interfered with by other phenomena, which irremediably generates imperfect data. In the literature, it is possible to find different typologies of imperfections of data (Parsons, 2001; Smets, 1997), but digital instrumentation is mainly affected by the following ones:

- the uncertainty characterizes the degree of conformity to reality;
- the inaccuracy is the lack of accuracy of measurement;

- the incompleteness is due to the incapacity of the device to measure the complete phenomenon or due to missing values;
- the inconsistency may appear in some specific conditions or in an array of sensors or multi-sensors devices when the various measurements conflict.

All these aspects can be approached with fuzzy logic in its broadest sense. Indeed, the possibilistic framework is made for modeling uncertainty. Inaccuracy is well handled by fuzzy numbers and more generally fuzzy sets. The incompleteness of data is not directly modeled with fuzzy logic, but most of the models (fuzzy decision trees, fuzzy rule bases) can be used with missing data for instance. Finally, the inconsistency may be difficult to extract from data. Nevertheless, expert knowledge can be used to detect, and sometimes correct, this kind of imperfection.

This is why we have been working for years to bring these communities together, at least within CEA.

6.2 Using knowledge to build interpretable features

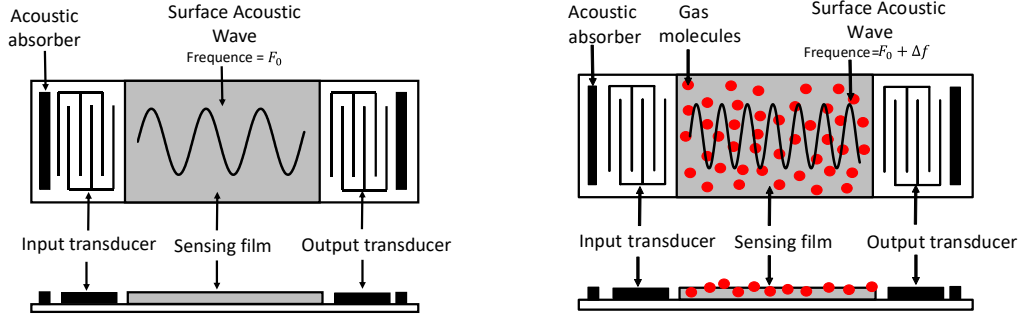
This section corresponds to Olivier Hotel's thesis work (Hotel, 2017) that we are continuing with Edwin Friedmann (for the ExpressIF[®] part), Christine Mer (for the sensor part) and many others (for the electronic part). One of the teams we are working with is developing chemical sensors based on diamond. In particular, we are interested in a technology based on SAW (Surface Acoustic Wave) to recognize chemical compounds.

6.2.1 SAW sensors and diamond functionalization

Broadly, chemical sensors transform chemical information into a measurable signal. These sensors are most often composed of a sensitive layer that transforms chemical information into a form of measurable energy and a transducer, which converts this energy into a measurable signal. Several sensor technologies allow converting the presence of a molecule on the sensitive layer into such a signal. In this work, we only consider SAW sensors.

These sensors are resonators made of a piezoelectric material on which are deposited two interdigital transducers (IDT) at the ends of a cavity covered with a sensitive layer. The IDTs generate a surface acoustic wave by converting electrical energy into mechanical energy. The system is designed in such a way that the cavity is resonant. The resonant frequency of the cavity is particularly sensitive to surface modifications. Thus, depending on the interactions between the sensitive layer deposited on the cavity and the surrounding environment, the measured resonant frequency will change. The physico-chemical interactions of the target molecules with this sensitive layer disturb the propagation of the acoustic wave on the surface of the piezoelectric resonator and thereby induce a measurable variation in frequency. The principle of SAW sensors is shown in figure 6.1. Regarding the direction of the propagation wave, we can distinguish Love-wave (in the surface plane) and Rayleigh-wave (in a plane orthogonal to the surface) SAW sensors. Our sensors were Rayleigh-wave SAW sensors.

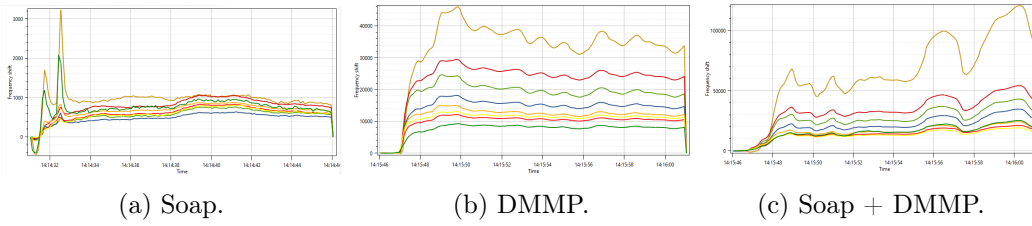
The sensitive layer of a SAW sensor is a key element to develop a multi-sensors approach to address the recognition of chemical signatures. It influences not only the selectivity of the sensor, i.e. its ability to respond only to certain target molecules, but also its sensitivity, i.e. its ability to respond to its target molecules even if these are at very low concentrations. The use of diamond as a sensitive layer is an original approach developed at CEA. Figure 6.2 shows a typical frequency shift due to the exposition of soap, DMMP (that is an explosive) and both to an array of 8 SAW sensors.



(a) A SAW sensor before any volatile chemical compound.

(b) A SAW sensor during an exposition to volatile chemical compounds.

Figure 6.1: Principle of Rayleigh-wave SAW sensors.



(a) Soap.

(b) DMMP.

(c) Soap + DMMP.

Figure 6.2: Examples of frequency shifts due to exposition to different compounds.

We now describe the physical phenomena that influence the resonant frequency in such sensors. Ballantine *et al.* (Ballantine et al., 1997) and Tard (Tard, 2013) have shown that the resonant frequency variation of SAW sensors is the superposition of three phenomena:

- A mass loading effect that is due to the increase in mass of the sensitive layer following the adsorption of molecules:

$$\Delta f_m = -\frac{C_m F_0^2}{\rho_S} \Delta m \quad (6.1)$$

where Δf_m is the frequency shift due to this effect, F_0 is the resonant frequency, C_m is called "mass sensitivity factor" and ρ_S is the surface mass density. Two essential elements are highlighted by this equation: first, SAW sensors can detect any kind of molecules since they all have a mass, and then the mass loading effect induces a negative frequency shift.

- A viscoelastic contribution that is due to the storage and the dissipation of power due to the deformation of the sensitive layer under the effect of the wave. The frequency shift is given by:

$$\Delta f_v = \frac{2\pi C_3}{v_0^2} \Delta \left(E \frac{4\nu - 5}{5\nu^2 + \nu - 4} \right) F_0^2 \quad (6.2)$$

where Δf_v is the frequency shift due to this contribution, E is the Young's modulus, ν the Poisson coefficient, C_3 is the surface velocity in one direction and v_0 the acoustic wave velocity.

- Finally, an electro-acoustic due to the coupling wave-charge carriers with the equation

$$\Delta f_e = -\frac{K^2}{2} \Delta \left(\frac{1}{1 + \frac{v_0 C_q}{\sigma_s}} \right) F_0 \quad (6.3)$$

where Δf_e is the frequency shift due to this contribution, σ_s is the surface conductivity, C_q denotes the sum of the dielectric permittivities of the area and the substrate:

$$C_q = \epsilon_0 + \epsilon_q \quad (6.4)$$

with K the electromechanical coupling coefficient. For sensors like ours, $K=0.11$.

The magnitude of the two first contributions is tens of Hertz while the last one is few micro-Hertz. Thus, the last contribution is negligible compared to the other two.

It has also been established that these disturbances can be modeled by first-order linear differential equations (Raj et al., 2012):

$$\begin{aligned} \tau_m \frac{\partial f_m}{\partial t} + f_m &= K_m c \\ \tau_v \frac{\partial f_v}{\partial t} + f_v &= K_v c \\ \tau_e \frac{\partial f_e}{\partial t} + f_e &= K_e c \end{aligned} \quad (6.5)$$

where c is the concentration profile, f_m, f_v, f_e are respectively the mass, visco-elastic and electro-acoustic contributions. The coefficients τ_m, τ_v and τ_e are the time constants: they characterize how quickly the frequency shift changes over time. The coefficients K_m, K_v and K_e are gains of these equations: they characterize the steady state of the solutions of these equations. The total frequency variation is then given by:

$$\begin{aligned} \Delta f &= f_m + f_v + f_e \\ \Delta f &\approx f_m + f_v. \end{aligned} \quad (6.6)$$

Perfectly selective sensors, that is to say, sensors that only respond to a single type of molecule, are rare. Regarding SAW sensors, the fact that they respond to any type of molecule, due to the mass contribution, makes them very unselective. To remedy this problem, it is common to use several sensors that are not very selective but that have different chemical affinities. The response of each sensor to a given molecule will therefore be different and will thus generate a chemical signature.

Our contribution consists in using this knowledge to build new features instead of the traditional ones (e.g. amplitudes of steady state, moments, ...).

6.2.2 Estimation of the mass and visco-elastic contributions parameters as new features

The equations of the previous sections are important in the sense that their constants τ_i, K_i with $i \in \{m, v\}$ can form a signature to one or several molecules. Our approach consists in solving this system:

$$\begin{cases} \tau_m \frac{\partial f_m}{\partial t} + f_m = K_m c \\ \tau_v \frac{\partial f_v}{\partial t} + f_v = K_v c \\ f_m + f_v = f. \end{cases} \quad (6.7)$$

Note that f is the frequency shift and it is measured by the sensor.

Unfortunately, to the best of our knowledge, it is not possible to solve analytically this system. We thus use a metaheuristic approach to solve it. We benchmarked several approaches and chose Particle Swarm Optimization (PSO) for its fast execution and its stability. The four constants K_m, τ_m, K_v, τ_v for each sensor allow solving different types of problems.

Beyond the volatile compounds recognition, which was the main application, with further manipulations of the previous equations whose solutions can be approximated with our approach, it is also possible to estimate the concentration profile with less than 10% error. This is a new application that was not possible before or at least that gave poorer results. Using the same principles, we designed also algorithms to select which functionalization was useful for a given application. We are not describing these methods here, but the interested reader can refer to Olivier Hotel’s PhD manuscript (Hotel, 2017).

6.2.3 Towards the recognition of mixture of compounds

The recognition of the elements in a mixture of compounds is a classical challenge with chemical sensors. The problem is very difficult because the mixture is made of an unknown number of gases, with different concentrations.

The first element identified as essential for identifying mixtures is the completeness of the training set, i.e. the presence of examples corresponding to all possible mixtures. Thus, two approaches can be formulated according to whether the acquisitions of the responses of the sensors to the mixtures of interest have been carried out or not. Let N_g and N_c denote respectively the number of gases and the number of concentrations of interest. There are $(N_c + 1)^{N_g} - 1$ possible combinations. Of course, having a method that handles automatically the mixtures from the base components would drop this number of combinations to $N_g N_c$.

The second element identified is the a priori knowledge of the number of compounds constituting a mixture: if it is known, it is then possible to reduce the number of possibilities. For example, if one wishes to identify a mixture of n compounds belonging to a set containing N_g , then the knowledge of N_g makes it possible to reduce the number of possibilities from $2^n - 1$ to $C_n^{N_g}$.

N_g	Complete training set	Partial training set
Known	Training of N_g classifiers Add N_g to the feature set	Open problem
Unknown	One class per mixture	Open problem

Table 6.1: Typology of problems and approaches for mixture recognition.

Table 6.1 shows a summary of the different approaches to deal with mixture of compounds. We thus formulated a regression problem to estimate the number of gases N_g in a mixture presented in (Hotel, Poli, Mer-Calfati, et al., 2017a).

6.2.4 Application

Our experiments showed that the K_i constants were sufficient to increase the performances of the main classifiers. Indeed, we experimented the approach on three datasets: 5 gases in laboratory with different concentrations, 21 kinds of coffee capsules, 4 compounds in a backpack.

Table 6.2 shows the results of the classification (accuracy) on the three datasets. The classification is performed by a Support Vector Machine (SVM), a Large Margin Nearest Neighbors (LMNN) and Bagged Trees (BT). The inputs are either the classical amplitudes or the constants K_m and K_v (our features). The results are given in percentages as a mean and a standard deviation and obtained after a 5-fold cross-validation. The results show that in almost all the cases, the classifiers based on our features outperform the ones based on amplitudes.

	Gases		Coffee		Backpack	
	Amplitudes	Constants K	Amplitudes	Constants K	Amplitudes	Constants K
SVM	92.7 ± 3.2	94.4 ± 5.1	62.0 ± 8.7	63.8 ± 8.4	96.6 ± 3.2	98.3 ± 1.1
LMNN	93.0 ± 2.8	95.8 ± 4.8	62.6 ± 8.7	69.8 ± 7.9	96.2 ± 2.9	96.6 ± 4.1
BT	93.5 ± 2.1	94.7 ± 3.1	55.8 ± 6.6	65.3 ± 6.9	96.6 ± 2.6	94.2 ± 3.7

Table 6.2: Accuracies of 3 classifiers used either with the amplitudes or our features (constants K_m , K_v) on the three datasets and after a 5-fold cross-validation.

Figure 6.3 shows the result of the concentration profile estimation process: the true concentration is 8 ppm (estimated 8.2 ppm) from 0 to 10 sec and 4 ppm (estimated 3.74) from 10 to 15 sec.

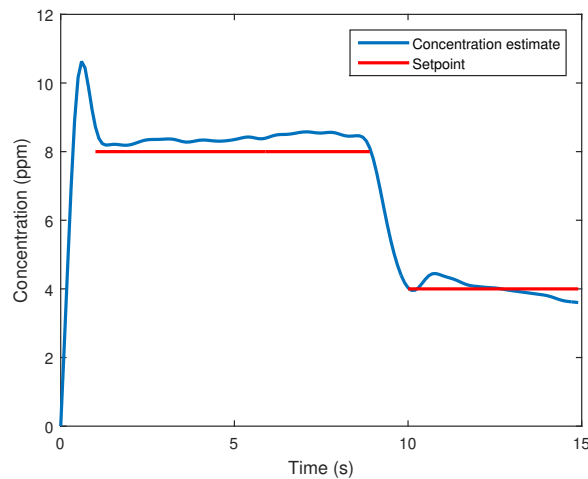


Figure 6.3: Concentration setpoint and concentration profile estimate.

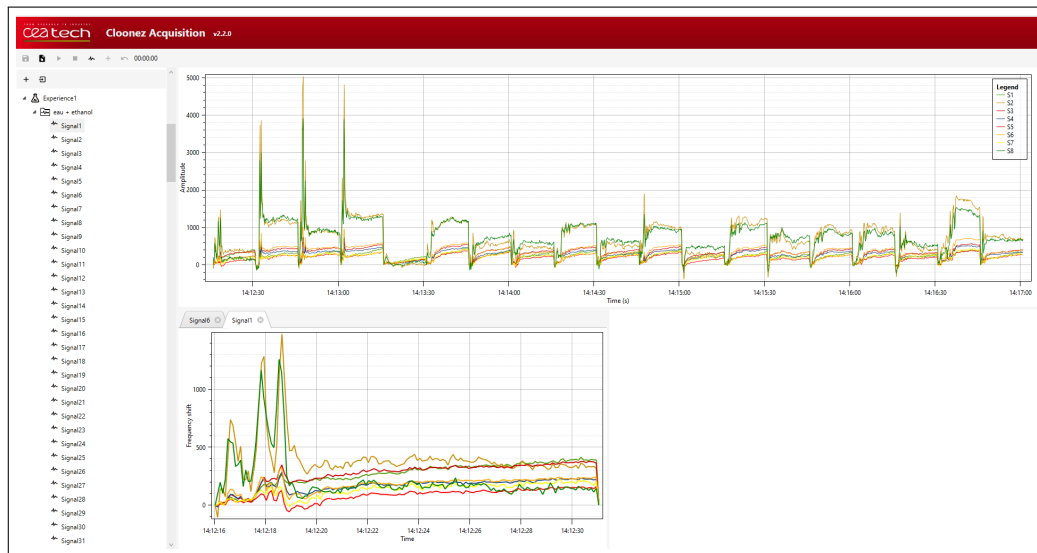


Figure 6.4: Screenshot of ClooNEZ software.

All the algorithms have been implemented in the CLOONEz software, which comprises three parts:

- The first part is dedicated to the acquisition of a properly labeled dataset (see figure 6.4). An experiment is made of several acquisitions of eventually different gases.

These acquisitions can be divided into signals that will be used for machine learning.

- The second part is a wizard that guides a user through the different steps to train and compare models. It allows selecting the features, the preprocessing and then the model. It allows fine-tuning hyperparameters with grid search and performing cross-validations.
- Finally, the last part allows connecting a sensor and loading a model to perform online analysis.

6.3 Adapting fuzzy decision trees to specific data

This work has been achieved by Arnaud Grivet Sébert under my supervision, within the European project CBord. When we collaborate with physicists, we often depend on their modeling of the phenomena. In physics, the evaluation of the uncertainty is of paramount importance, and they often model it using probabilistic distributions. So when physicists are in charge of preprocessing raw data, it is common to receive Gaussian distributions as input for our classifiers. Of course, it is questionable to use fuzzy logic because there are many frameworks (e.g. the possibilistic framework, etc.) (Destercke, 2022) that can be used to model this kind of uncertainty. Moreover the cohabitation between fuzzy logic and probabilities is often avoided but to qualify the probability of an event like in the proposition “the event is very probable”.

To process such data, there are mainly two options: either we use only the mean of the Gaussian distributions and usual classifiers, or we adapt classifiers to use the whole information contained in the Gaussian distribution. We investigated the second option. We chose to adapt two algorithms: a clustering method to obtain the terms of the linguistic variables, and the classical fuzzy decision tree induction algorithm. I first describe the context of this work.

6.3.1 Context

The rapid growth of transport activity is both the consequence and the support of the globalization of the economy. Ports and borders are experiencing an ever-increasing dynamic: for instance, Rotterdam Port, which is the first port in Europe, managed more than 8.6 million containers for 469 million tons in 2019 (Rotterdam Authority, 2022) (more recent figures are not representative due to COVID pandemic).

In parallel with such an increase in the transit of goods, it is necessary to adapt the security measures while the systematic inspection of the containers is no longer possible (Vesky, 2008). Most ports are equipped with x-ray scanners; however, they provide a limited vision of the content since the detection of risk is based solely on the shapes of the objects. If it is useful for the detection of weapons or illicit objects, the method remains ineffective for the detection of tobacco, drugs or explosives.

A new emerging non-invasive technique is based on tagged neutron technology (Perot et al., 2006; Pesente et al., 2004), which allows to pass through the walls of the container and to obtain the proportions of different atoms inside the examined voxel, i.e. a limited volume of the container.

A particle generator is used to produce a neutron beam through a fusion reaction in which an alpha particle is emitted almost back-to-back with the neutron. A first detector is used to localize the alpha particle to deduce the direction of the neutron. Gamma rays resulting from the interaction between emitted neutrons and materials in the container are detected using scintillators. Any voxel of the container can be examined thanks to the combination of the time difference between gamma and alpha detections and the neutron

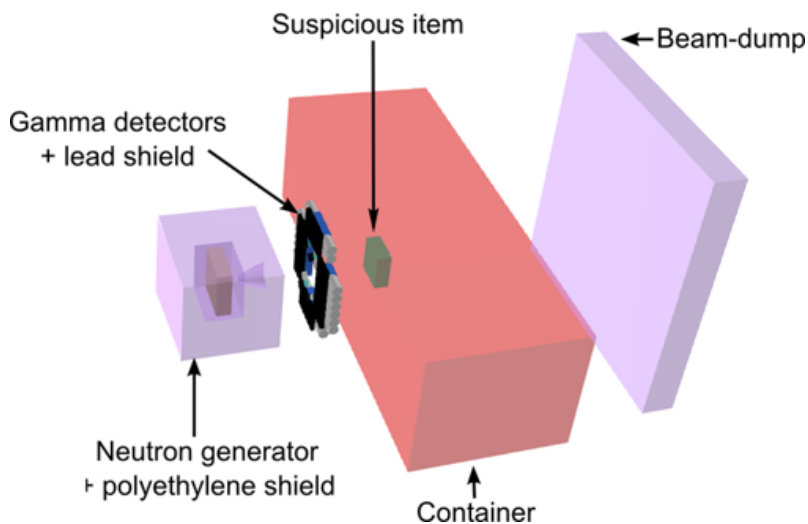


Figure 6.5: Overview of tagged neutron system in C-BORD project (Sardet et al., 2016).

direction. In practice, this voxel is typically selected after the x-ray image examination. Figure 6.5 represents the tagged neutron system of the C-BORD project.

The quantity of 19 pure elements is deduced through gamma-ray spectroscopy: the gamma energy spectrum is unfolded using a database of elementary gamma signatures induced by tagged neutrons, previously determined during calibration acquisitions.

The unfolding method consists in finding a linear regression model from the contributions of the elementary signatures: this is achieved by a non-negative least square (NNLS) optimization method (Lawson and Hanson, 1995). Unfortunately, the acquired data suffer from several imperfections: for instance, the particles interact with the material near the detectors, there is also an ambient natural radioactivity to consider, etc.

To assess the uncertainties related to the counting statistics, hundreds of synthetic spectra are generated from the original gamma spectrum of the chosen voxel: each channel of the spectrum is sampled according to a Poisson distribution. The NNLS procedure is applied to these hundreds of spectra to get a count fraction distribution of pure elements.

Formally, our imperfect data are represented by a vector of length N such as

$$\begin{bmatrix} \mathcal{N}(\mu_1, \sigma_1) \\ \vdots \\ \mathcal{N}(\mu_N, \sigma_N) \end{bmatrix} \quad (6.8)$$

where each $\mathcal{N}(\mu_i, \sigma_i)$, $i \in \llbracket 1; N \rrbracket$ is a Gaussian distribution.

Each dimension of this vector corresponds to an attribute of the dataset. Thus, the classification consists in attributing a label $c \in \mathcal{C}$ where \mathcal{C} is the set of all possible labels (classes) to such a vector.

The purist reader may be uncomfortable by the use of a Gaussian distribution on the bounded domain $[0; 1]$. This is tolerated and very frequent in the application field. So we have adapted our algorithms to this unusual usage. Considering the bounded domain instead of \mathbb{R} and considering the Gaussian distributions rather than other distributions allowed us to simplify the calculations but in no way compromises the generality of the approach. The next subsections present the method.

6.3.2 Learning vocabulary from uncertain data

To improve the interpretability and to fit better the data, the vocabulary is extracted beforehand. At least, the linguistic terms have the same semantics throughout the FDT.

On the contrary, if modalities were learned at different levels of the tree, and so on different subsets, we would have different meanings or different labels for potentially close terms.

To keep the construction of the partition simple, the shapes of the membership functions are imposed as triangular, except for the extreme terms, which are semi-trapezoids to consider the extreme values of the domain.

With the number of fuzzy terms being set, a clustering algorithm is performed on the input data for each feature. The centers of the clusters for a given feature are computed and constitute the tops of the triangular or trapezoidal terms. Since the fuzzy partition is forced to be strong, the tops of the triangles suffice to characterize the whole partition and the bases of the triangles are directly deduced. The terms are labeled manually after they have been learned.

The clustering algorithm to build the fuzzy partition is k-medoids (Kaufman and Rousseeuw, 1987) and its optimization PAM (Schubert and Rousseeuw, 2019). This well-known algorithm takes the number k of clusters as a parameter. It randomly chooses k objects from the dataset to be the initial centers of the clusters and the other objects are assigned to the closest center. In each cluster, the object minimizing the sum of the distances to the other objects of the cluster is set as the new center. Then the algorithm alternatively updates the sets belonging to each cluster and the centers until convergence. The optimizations suggested in (Schubert and Rousseeuw, 2019) prevent from the random initialization and make the execution of k-medoids efficient and deterministic.

We adapted the k-medoids algorithm to cluster the uncertain data by substituting the traditional Euclidean distance with the dissimilarity defined below.

Let x and y be two uncertain data (uncertain values of some feature from two different samples), f_x and f_y the densities of the associated distributions, a and b the bounds of the definition domain of the data. The dissimilarity used in our adapted k-medoids algorithm is inspired by the symmetric difference of two sets. Replacing the union (resp. intersection) by the maximum (resp. minimum) of two probability densities, we can consider the dissimilarity $d(x, y)$ defined by:

$$d(x, y) = 1 - \int_a^b \min(f_x, f_y).$$

It can be proved that this dissimilarity is a distance for continuous strictly positive functions whose integral on $[a; b]$ equals to 1 and thus, a fortiori, for normalized Gaussian probability distributions on $[a; b]$.

Now the modalities for each variable are known, we can induce the tree.

6.3.3 Fuzzy decision tree induction from uncertain data

The construction of the tree also requires an adaptation to our uncertain data. In (Duch, 2005) and (Tsang et al., 2009), the authors use integration techniques to handle relations between uncertain data and crisp terms. We generalized these techniques to the computation of the membership degrees of uncertain data to fuzzy terms. It enables us to define a membership degree (represented by a punctual value) of an uncertain datum to a fuzzy set, which will be classically handled in the building process. Given one of the attribute x_i of an input vector x , f_{x_i} the density of the distribution modeling x_i , V a fuzzy set characterized by its membership function μ_V and defined on $[a; b]$, the membership degree of x_i to the fuzzy set V is defined as:

$$\tilde{\mu}_V(x_i) = \int_a^b f_{x_i}(t)\mu_V(t)dt.$$

The membership degree of an uncertain sample x to a node n of the tree is defined as:

$$d_n(x) = \prod_{n' \in \text{path}(n)} \mu_{V_{n'}}(x)$$

where $V_{n'}$ is the fuzzy set associated with the node n' and $path(n)$ are the nodes on the path from the root to n , including n . For the sake of readability, we omit the necessary projection of x on the relevant dimension before the evaluation of the membership. By convention, the membership degree of a sample to the root is 1.

This definition of the membership degree to a node amounts to using the probabilistic t-norm, namely the product, to model the conjunction of the fuzzy terms. We also consider the "fuzzy frequency" of a class $c \in \mathcal{C}$ in the node n :

$$fr_{c/n} = \frac{\sum_{x \in \mathcal{X} \cap c} d_n(x)}{\sum_{x \in \mathcal{X}} d_n(x)}$$

and the "membership frequency" of the examples to the node n :

$$fr_n = \frac{\sum_{x \in \mathcal{X}} d_n(x)}{\sum_{m \in S} \sum_{x \in \mathcal{X}} d_m(x)}$$

where \mathcal{X} denotes the training set and S is the set of the sibling nodes of n , including n .

The fuzzy entropy, which is the criterion used to choose which feature will split a node n , can now be defined:

$$H(n) = - \sum_{n' \in children(n)} fr_{n'} \sum_{c \in \mathcal{C}} fr_{c/n'} \ln(fr_{c/n'})$$

where \mathcal{C} is the set of the classes of the problem.

This fuzzy entropy, inspired from (Peng and Flach, 2001), is a generalization of Shannon entropy. For every attribute x_i of the problem, $H(n)$ is the weighted mean of the fuzzy entropy measures in every potential child n' of n , each of these potential children being associated with one of the fuzzy terms partitioning the domain of the current feature. The weights are the membership frequencies of the potential children among their siblings. The feature that minimizes the entropy, or maximizes the entropy gain $G(n) = H(parent(N)) - H(n)$, $parent(N)$ being the father of n , is then used to split n , unless one of the following stopping criteria is true:

- all the features of the problem have been used on the path from the root to n ;
- there is not enough samples in n (fuzzily speaking, which means that the sum of the membership degrees of the training samples to n is less than a fixed threshold);
- splitting the node no longer sufficiently helps to discriminate the classes (mathematically translated by the fact that the entropy gain is less than a fixed threshold).

6.3.4 Decision making

Let x be a sample to classify. The membership degree of x to the leaves of the tree is computed in the same manner as for a training sample and with the same treatment of uncertainty. The weighted voting method (Ishibuchi, Nakashima, and Morisawa, 1999) is used to compute the confidence degree $\text{conf}(x \in c)$ of the proposition " x belongs to the class c ". For each class c , this confidence degree is the sum, on all the leaves L , of the product of the fuzzy frequency of c in the leaf with the membership degree of x to the leaf:

$$\text{conf}(x \in c) = \sum_{l \in L} fr_{c/l} \times d_l(x).$$

6.3.5 Empirical comparison

We generated a toy dataset, as close as possible to the real-world application, to compare two approaches: a regular fuzzy decision tree (that uses only the mean of the uncertain data) and our approach (that uses the whole distribution of the uncertain data).

The goal is to recognize materials from the uncertain proportions of atoms. We used the theoretical stoichiometric percentages of each element in each material (class), which provided us with reference values for each pair (class, element). To create an example x of a class c , a value m is randomly generated for each element e in an interval around the reference value of the pair (c, e) . The span of this interval is proportional to the reference value, the proportionality coefficient being called the degree of uncertainty in the following. The mean value of the Gaussian representing the uncertain proportion of e in x is set to m . A standard deviation is then randomly generated in an interval whose span is proportional to the one of the interval used to generate the mean. To generate the data, we used the chemical formula of seventeen explosives and nine drugs. Since the real data will be few due to the financial and temporal costs of the physical experiments, we chose to generate only ten examples per class to create the toy dataset.

We set the number of terms per linguistic variable according to a grid search. We considered 19 different atoms. C, N, O had respectively 5, 14, 12 terms, and the other atoms had 5 terms each. We considered 23 classes: 17 kinds of explosives and 6 kinds of drugs. We generated 10 datasets following the previous protocol and performed a 5-fold cross-validation.

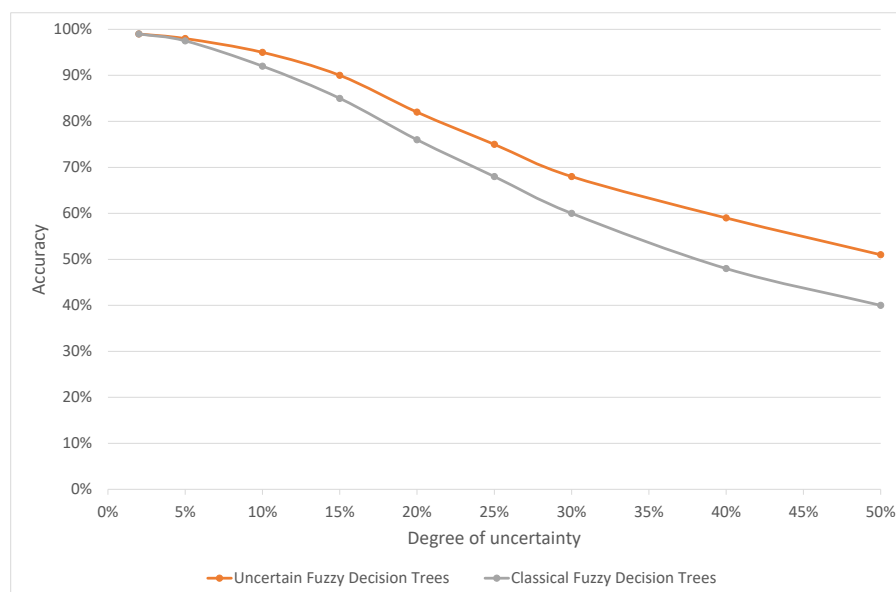
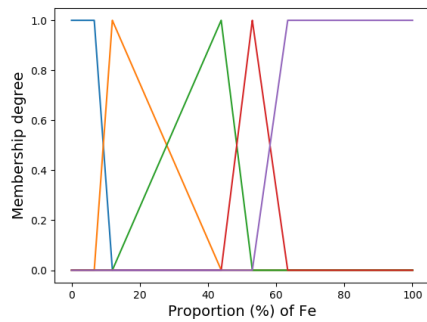


Figure 6.6: Impact of the degree of uncertainty on the accuracy on a toy dataset, for uncertain fuzzy decision trees and regular fuzzy decision trees.

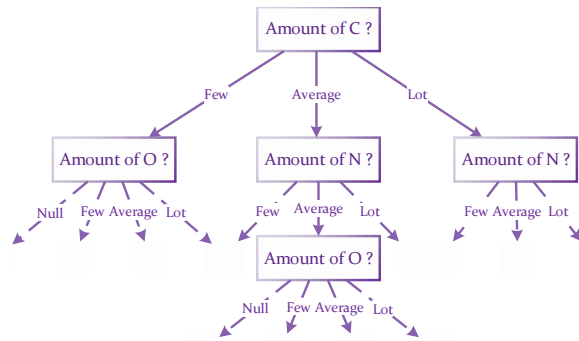
Figure 6.6 shows the average accuracy of both classifiers regarding the degree of uncertainty. The greater the uncertainty, the wider the gap. This experiment showed that it is important to consider the whole distribution since our approach performed always better than the regular fuzzy decision trees.

6.3.6 Application

To evaluate our method, we used two different real datasets: one from Saclay trials and one from Rotterdam trials. Due to the discrepancy between these two datasets, all of the results presented in the following come from five-fold-cross-validations, performing both training



(a) 5 terms automatically generated for Fe.



(b) Extract from a fuzzy decision tree for material classification.

Figure 6.7: Example of a fuzzy decision tree for material classification.

and evaluation on data from the same experiment (either Saclay's or Rotterdam's). For each of the two cases, we divided the dataset in such a way that, for every class c , there is the same proportion of samples belonging to class c in each of the five parts.

Classifying data by couple (matrix, target)

Contrary to what we expected, real samples contain two materials: the matrix material (the material that contains the target) and the target material (the suspicious content). This made the classification much more difficult because the classes, represented by the couples (matrix, target), were more numerous and the differences between the classes were more tenuous. Indeed, we had to consider 28 classes. Moreover, since more acquisitions were performed for some couples (matrix, target) than for others, the classes in the real datasets were unbalanced.

To address this problem, we removed some randomly chosen samples (undersampling) of the classes with more than ten samples. On the contrary, we created new samples (oversampling) of the classes with less than ten samples in the following way: a new sample of a class c was a randomly weighted convex combination of the existing samples of class c . A confusion matrix summarizing the results of a five-fold cross-validation on a balanced dataset from Saclay experiments is shown in Figure 6.8. The correct classification rate is equal to 72.5%.

This classification by couple (matrix, target) had a major drawback. New matrices or targets will not be recognized if they are not present in the training set. That is why another mode of classification is more relevant if unknown materials are to be analyzed.

Classifying data in four general classes and mixtures

Materials can be gathered into four main classes: ceramic, metallic, organic and other chemical materials. This last class involves materials containing elements that are neither ceramic, nor metallic, nor organic elements (potassium, sulfur, calcium, ...) – for instance mustard gas or potassium chloride. As mentioned above, most of the real samples contained a matrix and a target of two different materials, which might belong to two different classes among these four big classes. Thus, we might have samples labeled as "metallic + organic" for instance, and the algorithm should have been able to classify some samples as belonging to two classes. We had to perform a double-label classification. The order of the two classes was not taken into account since it was impossible to distinguish the matrix and the target with the input data. In the training phase, every double-labeled sample was artificially substituted by two single-labeled samples with the same features as the double-labeled one: e.g., a sample labeled "ceramic + organic" was replaced by a sample labeled "ceramic" and

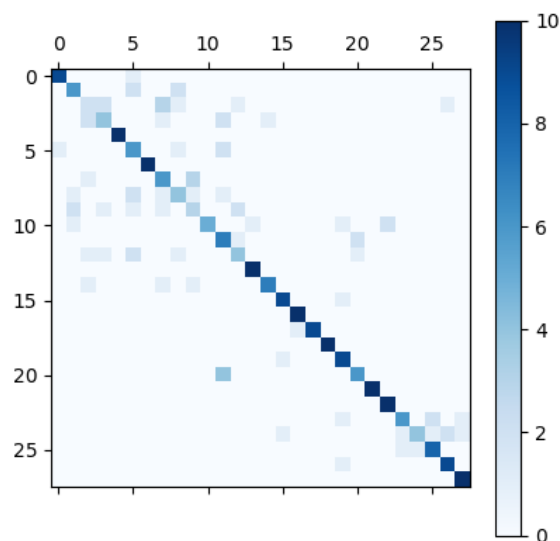


Figure 6.8: Confusion matrix from a cross-validation on balanced Saclay dataset for the (matrix, target) classification.

a sample labeled "organic". A double-labeled sample x would thus contribute to both of its classes: if x' was a new sample to recognize, similar to x in terms of features, and c_1 and c_2 were the two classes of x , the contribution of x would increase the scores $\text{conf}(x' \in c_1)$ and $\text{conf}(x' \in c_2)$. This technique allowed us to handle double-labeled classification without modifying the algorithm but only the training set. In the testing phase, we adapted the way of computing the correct classification rate. Let us consider a sample x for which $\text{conf}(x \in c)$ has been computed for every class c of the problem. If the true label of x was simple (not double), let us note c_0 its class. In this case, the classification of x was naturally considered correct if, and only if, the class c maximizing $\text{conf}(x \in c)$ was c_0 . Otherwise, if the true label of x was " $c_1 + c_2$ ", where c_1 and c_2 were two classes of the problem, x was considered to be well-classified if, and only if, the two classes with the highest confidence degrees were c_1 and c_2 , no matter the order. In the four-class problem, the classes were still unbalanced but the gap was less than for the (matrix, target) problem since all the classes had many samples. Therefore, we did not perform any balancing process. Figure 6.9 displays the confusion matrix of a five-fold cross-validation on Rotterdam data for the four-class problem. There is no ceramic materials in Rotterdam dataset. It explains why the problem has only six classes: "other chemicals", "metallic", "organic", "other chemicals + metallic", "other chemicals + organic", "metallic + organic". The classes are represented in this order in Figure 6.9.

The correct classification rate was very high – 95.5% – and the algorithm performed almost perfectly on all the classes, except "metallic + organic". Indeed, most of the samples of this class were classified as "other chemicals + metallic". This may be because typical elements of "other chemicals" class such as sulfur, chlorine or potassium were mistakenly present in the chemical proportions because of the unfolding process biases. Another probable reason was that chlorine was used in cocaine simulant. The correct classification rate remained very high in spite of this problem because the class "metallic + organic" represented only 34 out of 760 samples. If we balanced the dataset, the obtained accuracy was approximately 83%.

Classifying organic materials in three classes

The principal threats we were looking for were drugs and explosives, which are organic products. That is why we needed to perform another classification among organic mate-

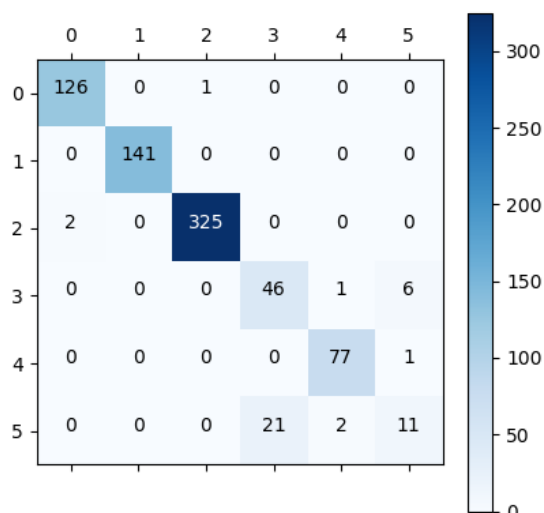


Figure 6.9: Confusion matrix from a cross-validation on Rotterdam data for the four-class and mixtures problem.

rials. The three classes we chose to distinguish are "drugs", "explosives" and "benign". Figure 6.10 shows the confusion matrix from a five-fold cross-validation on the organic Rotterdam data. In this dataset, there were 86 explosive samples, 85 drug samples and 268 benign samples. To cope with this unbalanced set, we randomly removed benign samples to keep only 80 benign samples. We did not perform oversampling for explosives or drugs as before because we already had a sufficient amount of data and oversampling might artificially increase the accuracy of the algorithm. The algorithm performed well at 78%, which was quite lower than for the four classes and mixtures problems. This three-class problem then seemed more difficult. Indeed, the classes, which were all organic, were more similar to each other than in the previous problem.

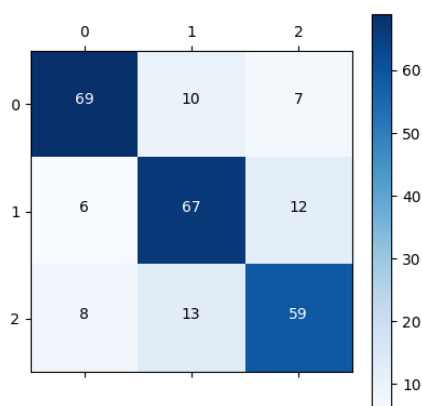


Figure 6.10: Confusion matrix from a cross-validation on Rotterdam data for the organic three-class problem.

If we considered two classes only (threat versus benign), the recall metric is equal to 0.89 and the precision is equal to 0.88 which were quite good results.

The complete classification process included a ceramic/metallic/organic/other chemicals classification and a further drug/explosive/benign classification for organic materials. The samples were finally classified among the six classes ceramic, metallic, other chemicals, drugs, explosives and benign organic, with the possibility of a mixture of two classes.

We studied the cause of classification errors. Some of these errors were due to the presence of unexpected pure elements in the composition. Another cause was the computation of uncertainties by the unfolding algorithm. For instance, we noticed an important drawback of this algorithm: anytime a proportion was equal to 0, the associated uncertainty was 0. That showed that the unfolding algorithm did not assess correctly some uncertainties.

6.4 Summary

We started this chapter with a plea about the use of fuzzy logic in digital instrumentation. Indeed, in our opinion, the framework is suitable to handle the different kinds of imperfections of sensors data. This motivates the development of ExpressIF Sensors[®] to reduce the gap between the sensors, which are the data source, and the fuzzy models to perform detection and recognition. We benefit from the fuzzy granules, extracted by well-known image and signal processing algorithms. The granules can then be used with words in a rule base, with a specific vocabulary regarding the type of data.

We then presented a first contribution that consists in creating more interpretable features by considering knowledge in physics. Indeed, the solutions of some specific equations allow for identifying the molecules interacting with the surface of a SAW sensor. These new features allow for tackling new problems like concentration estimation. The drawback of these features is that they are specific to this kind of sensors only. That work motivated the one presented in chapter 4, section 4.2: the automatic feature construction. Indeed, interpretable features led in our case to new applications and better performances in terms of classification. So automating their construction would make it possible to reproduce these results more generically. Step by step, we are replacing the usual classifiers by fuzzy classifiers, like FURIA for instance (Friedmann, Poli, et al., 2020).

Fuzzy rule-based models can bring some knowledge to the specialists in sensors. Indeed, in the case of the SAW sensors, since we have an array of sensors (we started with 8 sensors and we have now 32 sensors), we can see what sensors are important to classify the different volatile compounds. In addition, if a few sensors in the array are old or failing, we can immediately figure out the impact on the performances, telling eventually which classes will not be recognized anymore.

The second contribution of this chapter is an adaptation of existing algorithms to handle a specific type of data. Indeed, the data are usually scalars, but in the case of this work, each feature was a probabilistic distribution defined as a Gaussian, to represent the uncertainty of the value. We chose to adapt a clustering method to build the vocabulary and to adapt the entropy measure to induce a fuzzy decision tree. The results on a toy dataset, in which we could control the uncertainty degree, showed that handling the distribution performs better in terms of classification.

This chapter is the last one about the past contributions to improving fuzzy systems for XAI. We showed, through this part, we were able to tackle different kinds of problems: expressiveness, knowledge extraction, usability and also the direct link to the data sources.

Conclusion

We can only see a short distance
ahead, but we can see plenty
there that needs to be done.

Alan Turing

Chapter 7

Conclusion and perspectives

This document represents a retrospective of my research since my arrival at CEA. It is not exhaustive but the details can be found in the various publications and PhD manuscripts. The problem addressed consists in providing companies, hospitals and laboratories with tools for augmented decision-making. We follow the path of XAI, based on fuzzy systems and usable by end-users who are not necessarily in the field. This work has been carried out under constraints of funding and transferability, and led to the development of a software called ExpressIF[®], which is used by several companies and laboratories. It has also been motivated by real-world applications and needs. My contributions have made it possible to fill some gaps in fuzzy systems and explore new paradigms.

In terms of expressiveness, my first contribution is to have developed a fuzzy inference engine that can be extended in terms of functionalities, in particular in terms of the vocabulary that can be used in the knowledge base. This gave me the occasion to formalize different predicates in different domains (temporal, spatial, spatio-temporal, etc.). The predicates are assessable from the data, interpretable by a textual representation and seamlessly customizable.

I also explored new paradigms in machine learning under the constraints of interpretability and explainability. I defined those two terms that have different definitions through the literature. I compensated for the lack of representation learning in most of the transparent models by proposing automatic interpretable feature construction. I also proposed to extract knowledge from data. This knowledge provides both a comprehensive summary and can be executed during inference to classify, annotate, or predict. This knowledge can have the form of a conjunction of predicates or gradual rules. These approaches are directly inspired by human capacities, named by Zadeh "granulation", "organization" and "causation".

As ExpressIF[®] must be user-centric, I also contributed to the improvement of the usability of fuzzy systems. I tackled two antagonist problems: the knowledge edition and the explanations of the decisions. On the one hand, I worked on the ergonomics of the rule edition, to avoid the complexity of a small programming language. On the other hand, I worked on two case studies on the automatic generation of textual explanations. Both rely on the use of natural language generation tools. The first case study is about fuzzy classifiers and the second one uses insights from cognitive science and psychology to generate the explanation of semantic image annotation.

Finally, I wrote a plea for the use of fuzzy logic in digital instrumentation. I also carried out specific work on a chemical sensor and uncertain data from a tagged neutron instrument. I proposed to use knowledge to build more interpretable features that turned out to open up the possibilities in terms of applications: they allow not only to improve the chemical compound recognition but also to estimate the concentration profile or the number of gases in a mixture and to select which functionalization should be used in a given application. We also adapted existing algorithms for fuzzy decision tree induction

in order to handle uncertain data. In this particular case, the uncertainty was given as a Gaussian distribution.

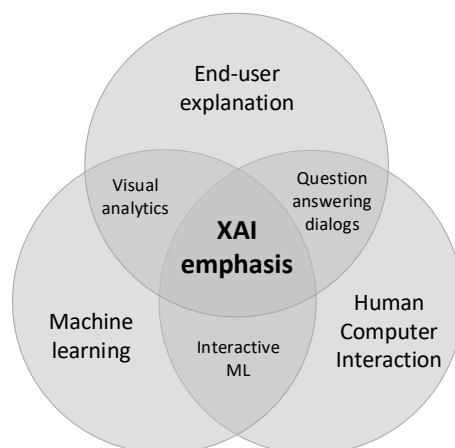


Figure 7.1: XAI emphasis (Gunning and Aha, 2019).

All these contributions fit well within the framework of XAI. As shown in figure 7.1, the XAI field requires efforts in several domains. As presented before, my contributions concern machine learning, end-user explanation, human-computer interaction, dialog and interactive ML.

I will present in the next sections the current and future work.

Current work

Causality and gradual rules extraction

Currently, I am still exploring the use of causality to build fuzzy rules. I collaborate with Aurore Lomet on different topics related to causality:

- We are working on time-to-failure prediction. We are designing a method based on causality to select relevant temporal relations between inputs and the time to failure. We then create rules based on the temporal predicates (see chapter 3, section 3.2) to recognize specific patterns in time series.
- We supervised Lucie Kunitomo-Jacquin during her postdoctoral fellowship. The goal is to adapt the causality extraction algorithms, like Peter-Clark (PC) and Fast Causal Inference (FCI), to fuzzy sets instead of classical variables.
- Since November 2022, we are supervising Antonin Arsac during his PhD, after his M2 internship, about causality in time series. Causality will serve as a selection algorithm to extract fuzzy rules for time series prediction.

I also supervised Killian Susini's internship. His goal was to finish the work on gradual rules. We compared three approaches in fuzzy logic to select the most suitable to predict monotonous functions: Gradual Generalized Modus Ponens (Vo, Detyniecki, and Bouchon-Meunier, 2013), Transformation-based Constraint-guided Generalized Modus Ponens (Lesot and Bouchon-Meunier, 2017) and an approach proposed by Dubois *et al.* (Dubois and Prade, 1992). We also replaced the genetic algorithm with a more robust method that can deal with few or big data, noisy or not. Some work is still needed to make the contributions clearer before publication.

Active learning

I am currently co-supervising Olivier Rousselle during his postdoctoral fellowship, with Nadia Ben Abdallah, who is a permanent researcher in the ExpressIF team. Olivier's topic concerns experimental data processing. We are exploring active learning to be able to suggest to a scientist the next experiment to be carried out. To achieve that, we need to meet some criteria:

- our method must propose an adaptive sampling, i.e. to decide sequentially the location of the next sample by balancing exploration and exploitation;
- we need to be able to incorporate scientific knowledge among the knowledge that has been automatically extracted from the experimental data;
- the method must be reliable and robust, i.e. a small change in the initial experiments should not lead to large changes in the predictions;
- our work falls within our global approach on interpretability and explainability;
- we must pay attention to performances (regarding both the prediction scores and the computation time).

To tackle this problem, we inspire from the Bayesian optimization framework. Basically, it consists in an iterative process that builds a predictive model (based on Bayesian process) and that maintain an uncertainty map to select the next experiment. In our case, for now, the model is a Sugeno rulebase. At each step, we learn both the vocabulary and the rules automatically.

Validation of a rule base

In terms of usability, I am working since the internship of Martin Everaert with Edwin Friedmann, on automatic checking of a rule base. The goal is to help the end-users writing a valid rule base, and provide them with advice on completeness and interpretability. The fact that the rule base is represented by a directed acyclic graph may help browsing the base for validation. For instance, it obviously helps finding cycles in the rule base. The difficulties emerge from the expressiveness of our system and we are developing methods based on the knowledge graph about predicates presented in chapter 3, section 3.5. It is also a real challenge because the algorithms that we can find in the literature are mainly combinatory and cannot be applied on real rulebases.

Semantic representation of an explanation

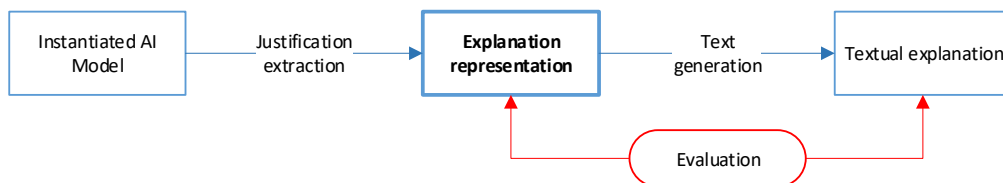


Figure 7.2: Overview of the proposed approach.

Finally, I work with Wassila Ouerdane (MICS, CentraleSupélec) and Titouan Lévêque (CentraleSupélec) on the representation of the content of an explanation. During Ismaïl Baaj's PhD, we investigated this problem. Figure 7.2 shows the overview of the approach we propose. Regarding our experience, we propose to separate the responsibilities: specialists of the AI model are in charge of extracting the justification and representing the

content of the explanation with our model. From this representation, specialists in Natural Language Generation (NLG) work on realizing the text. As a consequence, the evaluation of the automatic generation of textual explanations can be achieved in two separate parts: the evaluation of the representation and the evaluation of the realization. With Ismail, we explored the conceptual graph formalism as a candidate to represent the content of explanations. Nevertheless, in practice, it needs too many modifications to be used by NLG tools. We are thus investigating new formalisms that could be more suitable.

Future work

There are many topics that I want to explore with the ExpressIF team and other collaborations. The CEATech model implies that I need to find funds to carry out this research. I will highlight what kind of opportunities I am looking at to start these works.

Expressiveness

Expressiveness is still an asset of ExpressIF[®]. We continuously add new predicates regarding application domains. However, there are at least two domains we want to investigate. On the one hand, since we reason now on objects and regions of images, we need to study morphological predicates. In the spatial domain, such predicates allow describing shapes, like *IsElongated* predicate. In the spatio-temporal domain, these predicates will allow characterizing the changes of the shape, like *IsGrowing* predicate. These domains should have been studied during the DeepHealth project, notably for skin lesions. Unfortunately, we did not have a dataset representing the evolution of skin lesions. Morphological predicates can have direct applications:

- in crisis management, e.g. to make decisions based upon the spread of a fire, hazardous liquids, etc.;
- in ecology, e.g. to study the reduction (or hopefully the growth) of forest areas;
- in health, e.g. to characterize the growth (or the reduction) of a tumor.

On the other hand, given the positions of different entities, like humans in images from surveillance cameras, we would like to characterize group activities like the fact they come together and other interactions they may have. They can have direct applications in security, e.g. for the characterization of gang behaviors, but also to monitor a swarm of unmanned vehicles.

Laurence Boudet and I could collaborate on these two topics. This can be carried out by PhD candidates and then raised to a higher TRL during an European project or an industrial partnership.

Knowledge extraction

Knowledge extraction will remain a priority in the next years. The opposition "exploration versus exploitation" is also applying on my research.

In terms of exploitation, I have to go further into the approach based on relevant predicate learning. The expressiveness of the induced rule base is limited by the fact we only learn conjunctions of predicates. However, a predicate that is always false for a given class has also a discriminative power. Moreover, we still need to improve the efficiency of the method regarding the combinatorial factor. For now, I would like to collect more experiences on this new learning concept to prioritize the work on it. In a first time, collaborative projects and internships can be used to implement and experiment

this approach and then we can investigate the next development with a PhD co-supervised with Celine Hudelot (MICS, CentraleSupélec).

In terms of exploration, we have some opportunities to work on causality. In particular, we have to study the proprieties of such rules in terms of knowledge quality, interpretability and efficiency regarding the task (e.g. classification, regression, etc.). If we follow the paradigm proposed in chapter 4, section 4.4, we should be able to have different kinds of rules, not only gradual ones. A question will raise: how make those rules cohabit in the same rule base, in particular if they concern the same output? Collaborative projects are still needed to mature these ideas. In particular, I am trying to obtain internal projects related to material science, notably to fund a PhD that I will cosupervise with Nadia Ben Abdallah from ExpressIF team.

I also want to explore the continuous learning paradigm. This was the goal of Thomas Lamson's internship, but limited to Boolean logic, in which we tackled the notions of generalization and specialization. As I worked on learning with few data, it is possible that the algorithms over-generalize some concepts. With continuous learning, the algorithms may have to switch from generalization to specialization (or vice versa): this implies selecting the rules (or the piece of knowledge) that have to be changed and deciding how to change them. In addition, it is also important to be able to compare rules, to merge rule bases learned on different datasets. This work can be started during Olivier Rousselle's postdoctoral internship. Indeed, it could be useful not to learn the entire vocabulary set and rules at each iteration. We are considering with Nadia Ben-Abdallah to get other collaborative projects about material science to explore this topic.

Finally, I will start collaborating with Romaric Besançon from CEA List on the extraction of knowledge from scientific literature and patents, in particular in the material science field. We will use large language models to find causality in text, and more generally qualitative and quantitative information. For instance, some papers relate the causal relationship between some process parameters and some material performances. This can be confirmed by the automatic extraction of such relationships from the experimental data or it can complete them, before extracting rules. This collaboration will start in late 2023 within the ExIL (Extracting Information from Literature) project.

Usability

Usability is still important to us, but the actions we need to take are mainly technical. However, we will have to continue the efforts on the verification of rule bases and the generation of textual explanations.

For the first part, in the context of trustful AI, it is of paramount importance to be able to qualify a rule base. This relies at least on the completeness of the rule base, but we should be able to quantify how much a rule base is appropriate for a given task. We should benefit from the work on AI (or more generally software) certification. This is carried out within a collaboration with Edwin Friedmann from the ExpressIF team.

For the second part, the methods of textual explanation generation we proposed are based on simple cases. We need to expand the possible application cases by considering more and more ambiguities. Once the model of representation of explanation content will be effective, we will have to explore the adaptation of NLG methods. There is an aspect of explanations we have not explored yet: the human-machine interaction. Indeed, for now, the system provides the end-users with an explanation as a whole. We thus have to choose how verbose the system is: this can overflow the end-users if the explanations are too long, or, on contrary, let the end-users expect more details. We have to think about providing explanations as a dialog between the machine and the end-users. In my opinion, the formalism that will help modeling the content of the explanation should allow being queried to give different aspects of the explanation, notably with a hierarchical view in

terms of precision. For instance, we can start with a high-level explanation, and let the end-users ask for more details if needed. This is also what motivated the implementation of the dialog engine (see chapter 3, section 3.6.2) in an extensible way. Currently we are recruiting a new permanent researcher whose profile can help achieving this research. I consider to start a PhD in few years. Meanwhile, we are working with Edwin Friedmann, within an industrial project, on defining new feature importance algorithms that are specifically designed and optimized for fuzzy rule bases.

Digital instrumentation

The main perspectives in this domain are twofold. First, the representation and the consideration of the uncertainties of several sensors should use fuzzy logic. This can lead to several applications, notably isotopic recognition. For gamma imaging, which consists in superimposing gamma hot spots on the visible image, spatial predicates must be used to decrease the number of false hot spots. As a sensor detects a phenomenon within a specific area, spatial reasoning can be applied to "spatial measures" to help deciding where should be the next measurements or to deduce information from them.

Secondly, sensors should be able to learn autonomously from their locations. We can imagine applying federated learning to such scenarios where a global model could benefit from local models and vice-versa. The problem here is not a problem of confidentiality (even if it could be important), it is more a problem of background. For instance, different places have different natural nuclear background activities. It is almost impossible to travel everywhere in the world with the same sensor to acquire background activities. Thus, a decentralized approach can be important.

These problems can be tackled with the help of my colleagues from the department of digital instrumentation. As we did over the last years, we collaborate on European and industrial projects.

Extension to other problems solving

Currently, ExpressIF[®] has an inference engine and a fuzzy constraint satisfaction problem solver. As a perspective, we want to be able to tackle more combinatorial problems. Symbolic AI excels in this field and there are still needs for such problem solving in industry and Sciences.

In particular, we will address soon the problem of planning. This was not often addressed by the fuzzy logic community. We have to identify first where using fuzzy logic (in the broad sense) makes sense: should the states of the world, the plans or the temporal constraints be fuzzy? What is the impact on the computation time?

I am also interested in learning planning problems. It is somehow related to process mining that Laurence Boudet is exploring. Fuzzy logic may help capturing the inaccuracy of the observational data. We have first to explore ourselves this topic that is new for us before considering a PhD on it.

Towards the artificial research assistant

Finally, the field of material science gives us the opportunity to imagine an artificial research assistant, ExpressIF Materials[®]. Most of my contributions are already fitting this theme, in particular the ones presented in chapters 3, 4 and 5. This new platform aims at helping material science researchers in the discovery of new materials. This will be driven by various projects in that field, notably from the DIADEM (Integrated devices to Accelerate the Deployment of Emerging Materials) PEPR (Exploratory Priority Research Programme and Equipment).

In addition to all the previous perspectives that will nourish this project, I want to formalize through fuzzy logic the typical experimental research "trial and error" approach. This is yet another combinatorial problem that should be realized faster by AI. It is also related to continuous learning in the sense that the AI has to learn from the failures. A first step in that direction would be to design an experimental plan for the benefit of the training of the AI, as started during the post-doctoral fellowship of Olivier Rousselle.

References

- Achich, Nassira, Fatma Ghorbel, Fayçal Hamdi, Elisabeth Metais, and Faiez Gargouri (Sept. 2019). “A Typology of Temporal Data Imperfection”. In: *11th International Conference on Knowledge Engineering and Ontology Development (KEOD 2019)*. Vienna, Austria: SciTePress - Science and Technology Publications, pp. 305–311.
- Agrawal, R., T. Imieliński, and A. Swami (1993). “Mining association rules between sets of items in large databases”. In: *Acm sigmod record*. Vol. 22. 2, pp. 207–216.
- Alevizos, Elias, Anastasios Skarlatidis, Alexander Artikis, and Georgios Paliouras (2015). “Complex Event Processing Under Uncertainty: A Short Survey”. In: *Proceedings of the Workshops of the EDBT/ICDT 2015 Joint Conference*. Ed. by Peter M. Fischer, Gustavo Alonso, Marcelo Arenas, and Floris Geerts. Vol. 1330. CEUR Workshop Proceedings, pp. 97–103.
- Allen, James F. (Nov. 1983). “Maintaining Knowledge About Temporal Intervals”. In: *Communications of the ACM* 26.11, pp. 832–843. ISSN: 0001-0782.
- Alonso Moral, J.M., C. Castiello, L. Magdalena, and C. Mencar (2021). *Explainable Fuzzy Systems: Paving the Way from Interpretable Fuzzy Systems to Explainable AI Systems*. Studies in Computational Intelligence. Springer International Publishing.
- Alonso Moral, Jose M, Alejandro Ramos-Soto, Ehud Reiter, and Kees van Deemter (2017). “An exploratory study on the benefits of using natural language for explaining fuzzy rule-based systems”. In: *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 1–6.
- Arsac, Antonin, Aurore Lomet, and Jean-Philippe Poli (2023). *Causal discovery for time series with constraint-based model and PMIME measure*. arXiv: 2305.19695 [stat.ME].
- Artikis, A., C. Baber, P. Bizarro, C. Canudas-de-Wit, O. Etzion, F. Fournier, P. Goulart, A. Howes, J. Lygeros, G. Paliouras, A. Schuster, and I. Sharfman (2014). “Scalable Proactive Event-Driven Decision Making”. In: *Technology and Society Magazine, IEEE* 33.3, pp. 35–41. ISSN: 0278-0097.
- Aupetit, Michael, Ricardo De Aldama, Jean-Philippe Poli, and Laurence Boudet (2015). “Dispositif Et Procède D’optimisation D’un Systeme D’inference Floue Preservant L’interpretabilite”.
- Baader, F. (1996). “A Formal Definition for the Expressive Power of Terminological Knowledge Representation Languages”. In: *Journal of Logic and Computation* 6.1, pp. 33–54.
- Baaj, Ismaïl, Jean-Philippe Poli, Wassila Ouerdane, and Nicolas Maudet (2021a). “Inférence min-max pour un système à base de règles possibilistes”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- (2021b). “Min-max inference for Possibilistic Rule-Based System”. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Baaj, Ismaïl and Jean-Philippe Poli (2019). “Natural Language Generation of Explanations of Fuzzy Inference Decisions”. In: *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Baaj, Ismaïl, Jean-Philippe Poli, and Wassila Ouerdane (2019). “Some Insights Towards a Unified Semantic Representation of Explanation for eXplainable Artificial Intelligence

- (XAI)”. In: *1st Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence*, pp. 14–19.
- Baalouch, Marouen, Jean-Philippe Poli, Maxime Defurne, and Noëlie Cherrier (2019). “Sim-to-Real Domain Adaptation For High Energy Physics”. In: *ML for Physical Sciences at NeurIPS*.
- Ballantine, B., R. White, S. Martin, J. Ricco, E. Zellers, G. Frye, and H. Wohltjen (1997). *Acoustic wave sensors Theory, Design and Physico-Chemical Applications*. Academic Press.
- Barredo-Arrieta, Alejandro, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera (2020). “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”. In: *Information Fusion* 58, pp. 82–115. ISSN: 1566-2535.
- Barro, S., A. Bugarién, P. Cariñena, F. Diéaz-Hermida, and M. Mucientes (2008). “Fuzzy Temporal Rule-Based Systems: New Challenges”. In: *Actas del XIV Congreso Español sobre Tecnologías y Lógica Fuzzy (ESTYLF)*. Langreo (Spain), pp. 507–514.
- Belohlávek, R. (2012). *Fuzzy relational systems: foundations and principles*. Vol. 20.
- Biederman, I. (1981). *On the Semantics of a Glance at a Scene*.
- Bloch, I. (1999a). “On fuzzy distances and their use in image processing under imprecision”. In: *Pattern Recognition* 32.11, pp. 1873–1895.
- (2005). “Fuzzy spatial relationships for image processing and interpretation: a review”. In: *Image and Vision Computing* 23.2, pp. 89–110.
- Bloch, Isabelle (1999b). “Fuzzy Relative Position between Objects in Image Processing: a Morphological Approach”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21.7, pp. 657–664.
- Bloch, Isabelle and Henri Maitre (1995). “Fuzzy Mathematical Morphologies: A Comparative Study”. In: *Pattern Recognition* 28.9, pp. 1341–1387.
- Bloch Isabelle, Ralescu Anca (2023). *Fuzzy Sets Methods in Image Processing and Understanding: Medical Imaging Applications*. Springer.
- Bobrow, D. G., S. Mittal, and M. J. Stefik (Sept. 1986). “Expert Systems: Perils and Promise”. In: *Commun. ACM* 29.9, pp. 880–894.
- Borgida, A. (1996). “On the relative expressiveness of description logics and predicate logics”. In: *Artificial Intelligence* 82.1, pp. 353–367.
- Borji, Ali, Ming-Ming Cheng, Huaizu Jiang, and Jia Li (2014). “Salient object detection: A survey”. In: *Computational Visual Media* 5, pp. 117–150.
- Borst, Gregoire, Stephen Kosslyn, and Michel Denis (May 2006). “Different cognitive processes in two image-scanning paradigms”. In: *Memory & cognition* 34, pp. 475–90.
- Bouchon-Meunier, Bernadette (2007). *La logique floue*. Presses Universitaires de France.
- (2021). “Is Time Fuzzy?” In: *Statistical and Fuzzy Approaches to Data Processing, with Applications to Econometrics and Other Areas: In Honor of Hung T. Nguyen’s 75th Birthday*. Ed. by Vladik Kreinovich. Cham: Springer International Publishing, pp. 47–54.
- Bouchon-Meunier, Bernadette, Marie-Jeanne Lesot, and Christophe Marsala (Apr. 2021). “Lotfi A. Zadeh, the visionary in Explainable Artificial Intelligence”. In: *TWMS J. Pure Applied Math* 12.1, pp. 5–13.
- Boudergui, K, F Carrel, T Domenech, N Guenard, J-P Poli, A Ravet, V Schoepff, and R Woo (2011). “Development of a Drone Equipped with Optimized Sensors for Nuclear and Radiological Risk Characterization”. In: *2011 2nd International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and Their Applications*.
- Budescu, David V, Han-Hui Por, and Stephen B Broomell (2012). “Effective communication of uncertainty in the IPCC reports”. In: *Climatic change* 113.2, pp. 181–200.

- Cariñena, P., A. Bugarién, M. Mucientes, and S. Barro (2000). “A Language for Expressing Fuzzy Temporal Rules”. In: *Mathware and Soft Computing* 7.2-3, pp. 213–227.
- Cherrier, Noëlie (2021). “Interpretable machine learning for CLAS12 data analysis”. PhD thesis.
- Cherrier, Noëlie, Maxime Defurne, Jean-Philippe Poli, and Franck Sabatié (2019). “Embedded Constrained Feature Construction for High-Energy Physics Data Classification”. In: *ML for Physical Sciences at NeurIPS*.
- Cherrier, Noëlie, Michael Mayo, Jean-Philippe Poli, Maxime Defurne, and Franck Sabatié (2020). “Interpretable Machine Learning with Bitonic Generalized Additive Models and Automatic Feature Construction”. In: *Discovery Science*. Ed. by Annalisa Appice, Grigorios Tsoumakas, Yannis Manolopoulos, and Stan Matwin. Cham: Springer International Publishing, pp. 386–402.
- Cherrier, Noëlie, Jean-Philippe Poli, Maxime Defurne, and Franck Sabatié (2019). “Consistent Feature Construction with Constrained Genetic Programming for Experimental Physics”. In: *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, pp. 1650–1658.
- (2020). “Embedded Feature Construction in Fuzzy Decision Tree Induction for High Energy Physics Classification”. In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 615–622.
- Christiaens, Guillaume (2021). “Beam-spin asymmetry of deeply virtual Compton scattering off the proton at 10.6 GeV with CLAS12 at Jefferson Laboratory”. PhD thesis. University of Glasgow.
- Cinbis, R. G. and S. Aksoy (2007). “Relative Position-Based Spatial Relationships using Mathematical Morphology”. In: *2007 IEEE International Conference on Image Processing*. Vol. 2, pp. 97–100.
- Cohen, N., O. Sharir, and A. Shashua (2016). “On the Expressive Power of Deep Learning: A Tensor Analysis”. In: *29th Annual Conference on Learning Theory*. Vol. 49, pp. 698–728.
- Cohen, William W. (1995). “Fast Effective Rule Induction”. In: *Machine Learning Proceedings 1995*. Ed. by Armand Prieditis and Stuart Russell. San Francisco (CA): Morgan Kaufmann, pp. 115–123.
- Colliot, O. (2003). “Représentation, évaluation et utilisation de relations spatiales pour l’interprétation d’images. Application à la reconnaissance de structures anatomiques en imagerie médicale”. PhD thesis. Télécom ParisTech.
- Denis, Michel, Marie-Rose Goncalves, and Daniel Memmi (1995). “Mental scanning of visual images generated from verbal descriptions: Towards a model of image accuracy”. In: *Neuropsychologia* 33.11. The Neuropsychology of Mental Imagery, pp. 1511–1530. ISSN: 0028-3932.
- Destercke, Sébastien (Aug. 2022). “Uncertain data in learning: challenges and opportunities”. In: *Proceedings of the Eleventh Symposium on Conformal and Probabilistic Prediction with Applications*. Ed. by Ulf Johansson, Henrik Boström, Khuong An Nguyen, Zhiyuan Luo, and Lars Carlsson. Vol. 179. Proceedings of Machine Learning Research. PMLR, pp. 322–332.
- Dev Gupta, Sharmi, Begum Genc, and Barry O’Sullivan (Aug. 2021). “Explanation in Constraint Satisfaction: A Survey”. In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Ed. by Zhi-Hua Zhou. International Joint Conferences on Artificial Intelligence Organization, pp. 4400–4407.
- Di Jorio, Lisa, Anne Laurent, and Maguelonne Teisseire (2008). “Fast extraction of gradual association rules: A heuristic based method”. In: *Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology*, pp. 205–210.
- Doshi-Velez, Finale and Been Kim (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. URL: <https://arxiv.org/abs/1702.08608>.

- Dubois, D., H. Fargier, and H. Prade (1996). “Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty”. In: *Applied Intelligence* 6.4, pp. 287–309.
- Dubois, Didier and Henry Prade (1992). “Gradual inference Rules in Approximate Reasoning”. In: *The Computer Journal* 27.2, pp. 97–111.
- Duch, Wlodzislaw (2005). “Uncertainty of data, fuzzy membership functions, and multi-layer perceptrons”. In: *IEEE transactions on neural networks* 16.1, pp. 10–23.
- Farah, Martha J. and Stephen M. Kosslyn (1981). “Structure and strategy in image generation”. In: *Cognitive Science* 5.4, pp. 371–383. ISSN: 0364-0213.
- Freeman, John (1975). “The Modelling of Spatial Relations”. In: *Computer Graphics and Image Processing* 4.2, pp. 156–171.
- Friedmann, Edwin and Jean-Philippe Poli (2019). “Système à base de règles floues pour la reconnaissance de composés chimiques”. In: *LFA 2019 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Friedmann, Edwin, Jean-Philippe Poli, Olivier Hotel, and Christine Mer-Calfati (2020). “Fuzzy Classifiers for Chemical Compound Recognition from SAW Sensors Signals”. In: *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 917–922.
- Gatt, Albert and Ehud Reiter (2009). “SimpleNLG: A realisation engine for practical applications”. In: *Proceedings of the 12th European Workshop on Natural Language Generation*. Association for Computational Linguistics, pp. 90–93.
- Geurts, P. (2001). “Pattern Extraction for Time Series Classification”. In: *Principles of Data Mining and Knowledge Discovery*, pp. 115–127.
- Grivet Sébert, Arnaud and Jean-Philippe Poli (2017). “Vers la Classification de Matériaux à Partir de Compositions Chimiques Incertaines”. In: *LFA 2017 - Rencontres francophones sur la Logique Floue et ses Applications*.
- (2018a). “Fuzzy Rule Learning for Material Classification from Imprecise Data”. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, pp. 62–73.
- (2018b). “Material Classification from Imprecise Chemical Composition: Probabilistic vs Possibilistic Approach”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Gunning, David and David Aha (June 2019). “DARPA’s Explainable Artificial Intelligence (XAI) Program”. In: *AI Magazine* 40.2, pp. 44–58.
- Hajri, Hiba, Jean-Philippe Poli, and Laurence Boudet (2021a). “Towards Monotonous Functions Approximation from Few Data With Gradual Generalized Modus Ponens: Application to Materials Science”. In: *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 796–800.
- (2021b). “Vers l’approximation de fonctions monotones avec le modus ponens généralisé graduel à partir de peu de données : Application à la Science des Matériaux”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Hardaway, Donald E. and Richard P. Willi (1990). “A Review of Barriers to Expert System Diffusion”. In: *Proceedings of the 1990 ACM SIGBDP Conference on Trends and Directions in Expert Systems*. SIGBDP ’90. Association for Computing Machinery, pp. 619–639.
- Hendricks, L. A., Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell (2016). “Generating visual explanations”. In: *European Conference on Computer Vision*, pp. 3–19.
- Hotel, Olivier (Dec. 2017). “Algorithmes, méthodes et modèles pour l’application des capteurs à ondes acoustiques de surface à la reconnaissance de signatures de composés chimiques”. PhD thesis. Université Pierre et Marie Curie - Paris VI.

- Hotel, Olivier, Jean-Philippe Poli, Christine Mer-Calfati, Emmanuel Scorsone, and Samuel Saada (2017a). “Estimation of the Number of Volatile Compounds in Simple Mixtures”. In: *Proceedings of EuroSensors 2017*. Vol. 1. 4, p. 623.
- (2017b). “Estimation of the Parameters of Saw Sensor’s Frequency Shift: Application to Odour Recognition and Concentration Evaluation”. In: *2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*. IEEE, pp. 1–3.
- (2017c). “Saw Sensor’s Frequency Shift Characterization for Odor Recognition and Concentration Estimation”. In: *IEEE Sensors Journal* 17.21, pp. 7011–7018.
- (2018). “A Review Of Algorithms For Saw Sensors E-nose Based Volatile Compound Identification”. In: *Sensors and Actuators B: Chemical* 255, pp. 2472–2482.
- Hotel, Olivier, Jean-Philippe Poli, and Samuel Saada (2017). “Apports De L’optimisation Par Essaims Particulaires Pour L’identification De Composés Chimiques”. In: *18ème Conférence ROADEF de la Société Française de Recherche Opérationnelle et Aide à la Décision*.
- Hudelot, Céline, Jamal Atif, and Isabelle Bloch (2008). “Fuzzy spatial relation ontology for image interpretation”. In: *Fuzzy Sets and Systems* 159, pp. 1929–1951.
- Hühn, Jens and Eyke Hüllermeier (Dec. 2009a). “FURIA: an algorithm for unordered fuzzy rule induction”. en. In: *Data Mining and Knowledge Discovery* 19.3, pp. 293–319.
- (2009b). “FURIA: an algorithm for unordered fuzzy rule induction”. In: *Data Mining and Knowledge Discovery* 19.3, pp. 293–319.
- “Inductive Learning” (n.d.). In: *Advanced Artificial Intelligence*. Chap. Chapter 7, pp. 247–297.
- Iphar, Clément, Laurence Boudet, and Jean-Philippe Poli (2021a). “Evaluation d’une zone d’écoulement floue basée sur la topographie à l’aide de prédicats spatiaux 3D”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- (2021b). “Topography-based Fuzzy Assessment of Runoff Area with 3D Spatial Relations”. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- (2023a). “Evaluation floue d’une zone de feu basée sur la topographie pour la simulation de la propagation de feux de forêt”. In: *LFA 2023 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- (2023b). “Topography-based Fuzzy Assessment of Burning Area in Wildfire Spread Simulation”. In: *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*.
- Ishibuchi, Hisao, Tomoharu Nakashima, and Takehiko Morisawa (1999). “Voting in fuzzy rule-based systems for pattern classification problems”. In: *Fuzzy sets and systems* 103.2, pp. 223–238.
- Jacquin, Lucie, Aurore Lomet, and Jean-Philippe Poli (2021). “Discussion Sur La Causalité Imparfaite Et Son Application Aux Sciences Expérimentales”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Jimenez-del-Toro, O., H. Müller, M. Krenn, K. Gruenberg, A. A. Taha, M. Winterstein, I. Eggel, A. Foncubierta-Rodriéguez, O. Goksel, A. Jakab, et al. (2016). “Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks”. In: *IEEE transactions on medical imaging* 35.11, pp. 2459–2475.
- Kaufman, Leonard and Peter J Rousseeuw (1987). “Clustering by means of medoids. Statistical Data Analysis based on the L1 Norm”. In: *Y. Dodge, Ed*, pp. 405–416.
- Kellogg, R. T. (1980). “Feature frequency and hypothesis testing in the acquisition of rule-governed concepts”. In: *Memory & Cognition* 8.3, pp. 297–303.
- Kelly, Sage, Sherrie-Anne Kaye, and Oscar Oviedo-Trespalacios (2023). “What factors contribute to the acceptance of artificial intelligence? A systematic review”. In: *Telematics and Informatics* 77, p. 101925.

- Kohavi, Ron, George H John, et al. (1997). “Wrappers for feature subset selection”. In: *Artificial intelligence* 97.1-2, pp. 273–324.
- Kosslyn, Stephen M., Thomas M. Ball, and Brian J. Reiser (Feb. 1978). “Visual images preserve metric spatial information: Evidence from studies of image scanning”. English (US). In: *Journal of Experimental Psychology: Human Perception and Performance* 4.1, pp. 47–60. ISSN: 0096-1523.
- Koza, John R (1992). *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. MIT press.
- Kunitomo-Jacquín, Lucie, Aurore Lomet, and Jean-Philippe Poli (2022). “Causal discovery for fuzzy rule learning”. In: *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Langs, G., H. Müller, B. H. Menze, and A. Hanbury (2013). “VISCERAL: Towards Large Data in Medical Imaging - Challenges and Directions”. In: *MCBR-CDS MICCAI workshop*. Vol. 7723.
- Laurent, Anne, Marie-Jeanne Lesot, and Maria Rifqi (2009). “GRAANK: Exploiting rank correlations for extracting gradual itemsets”. In: *International Conference on Flexible Query Answering Systems*. Springer, pp. 382–393.
- Lawson, Charles L. and Richard J. Hanson (1995). *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics.
- Le Yaouanc, Jean-Marie and Jean-Philippe Poli (2012). “A Fuzzy Spatio-temporal-based Approach for Activity Recognition”. In: *International Conference on Conceptual Modeling*. Springer, pp. 314–323.
- Legendre, George L (2011). “Pasta by design”. In: *Architectural Design* 81.4, pp. 100–101.
- Lesot, M-J., M. Rifqi, and B. Bouchon-Meunier (2008). “Fuzzy prototypes: From a cognitive view to a machine learning principle”. In: *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, pp. 431–452.
- Lesot, Marie-Jeanne and Bernadette Bouchon-Meunier (July 2017). “A Local Transformation-based Constraint-guided GMP Compatible with an Interpolation Scheme”. In: *2017 International Conference on Fuzzy Systems*. Naples, Italy.
- Levesque, H. J. and R. J. Brachman (1987). “Expressiveness and tractability in knowledge representation and reasoning 1”. In: *Computational intelligence* 3.1, pp. 78–93.
- Lipton, Zachary C. (June 2018). “The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery.” In: *Queue* 16.3, pp. 31–57.
- Luccio, Riccardo (1993). “Gestalt problems in cognitive psychology: Field theory, invariance and auto-organisation”. In: *Intelligent Perceptual Systems: New Directions in Computational Perception*. Ed. by Vito Roberto. Springer Berlin Heidelberg, pp. 1–19.
- Magdalena, L. (2015). “Fuzzy Rule-Based Systems”. In: *Springer Handbook of Computational Intelligence*. Springer Berlin Heidelberg, pp. 203–218.
- Magne, Sylvain, Wilfrid Husson, Grigaut-Desbrosses Hans Payet Lucas, Guillaume Sutra, Stéphane Dogny, Pierre-Guy Allineï, Marion Ledieu, Olivier Guéton, and Bernard Leibovici (2021). “1D OSL/FO Dosimeter Array for Remote Radiological Investigations in Hard-to-access Zones”. In: *The 7th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications*.
- Mamdani, Ebrahim H. (1974). “Applications of fuzzy algorithms for control of a simple dynamic plant”. In: *Proceedings of the IEEE*.
- Marsala, C. and M. Rifqi (2017). “Fuzzy decision tree and fuzzy gradual decision tree: Application to job satisfaction”. In: *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–6.
- Miller, George A. (Mar. 1956). “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information”. In: *The Psychological Review* 63.2, pp. 81–97.

- Miller, Tim (2019). “Explanation in artificial intelligence: Insights from the social sciences”. In: *Artificial Intelligence* 267, pp. 1–38. ISSN: 0004-3702.
- Moore, Johanna (1994). “Assessment of Explanation Systems”. In: *Technology Assessment in Software Applications*.
- Moore, Johanna and William Swartout (Dec. 1989). *Explanation in expert systems: A survey*. Tech. rep. University of Southern California: Information Sciences Institute.
- Noton, David and Lawrence Stark (1971). “Scanpaths in Eye Movements during Pattern Perception”. In: *Science* 171.3968, pp. 308–311. ISSN: 0036-8075.
- Parsons, Simon (July 2001). “All about uncertainty”. In: *Qualitative Methods for Reasoning under Uncertainty*. The MIT Press. Chap. 2.
- Pasquier, N., Y. Bastide, R. Taouil, L. Lakhal, et al. (1999). “Efficient mining of association rules using closed itemset lattices”. In: *Information systems* 24.1, pp. 25–46.
- Peng, Yonghong and P Flach (2001). “Soft discretization to enhance the continuous decision tree induction”. In: *Integrating Aspects of Data Mining, Decision Support and Meta-Learning* 1.34, pp. 109–118.
- Perot, Bertrand, Gregory Perret, Alain Mariani, Jean-Luc Ma, Jean-Louis Szabo, Emmanuel Mercier, Guillaume Sannie, Giuseppe Viesti, Giancarlo Nebbia, Silvia Pesente, Marcello Lunardon, Paola Formisano, Sandra Moretto, Daniela Fabris, Aldo Zenoni, Germano Bonomi, Antonietta Donzella, Andrea Fontana, Gaia Boghen, Vladivoj Valkovic, Davorin Sudac, Marek Moszynski, Tadeusz Batsch, Michal Gierlik, Dariusz Wolski, Wlodzimierz Klamra, Patrick Isaksson, Philippe Le Tourneur, Miguel Lhuissier, Annamaria Colonna, Carlo Tintori, Paolo Peerani, Vitor Sequeira, and Martino Salvato (Apr. 2006). “The EURITRACK project: development of a tagged neutron inspection system for cargo containers”. In: *SPIE Defense and Security Symposium*. Vol. 6213. Orlando, United States, p. 621305.
- Pesente, Silvia, Giancarlo Nebbia, Marcello Lunardon, Giuseppe Viesti, Davorin Sudac, Karlo Nađ, Sasha Blagus, and Vladivoj Valković (2004). “Detection of hidden explosives by using tagged neutron beams with sub-nanosecond time resolution”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 531.3, pp. 657–667.
- Pierrard, Régis (2020). “Explainable Classification and Annotation through Relation Learning and Reasoning”. Thèse de doctorat dirigée par Hudelot, Céline Informatique université Paris-Saclay 2020. PhD thesis.
- Pierrard, Régis, Laurent Cabaret, Jean-Philippe Poli, and Céline Hudelot (2020). “SIMD-Based Exact Parallel Fuzzy Dilation Operator for Fast Computing of Fuzzy Spatial Relations”. In: *Proceedings of the 2020 Sixth Workshop on Programming Models for SIMD/Vector Processing*. WPMVP’20. Association for Computing Machinery.
- Pierrard, Régis, Jean-Philippe Poli, and Céline Hudelot (2018a). “A Fuzzy Close Algorithm for Mining Fuzzy Association Rules”. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, pp. 88–99.
- (2018b). “Learning Fuzzy Relations and Properties for Explainable Artificial Intelligence”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 1–8.
- (2021). “Spatial relation learning for explainable image classification and annotation in critical applications”. In: *Artificial Intelligence* 292, p. 103434. ISSN: 0004-3702.
- Pierrard, Régis, Jean-Philippe Poli, and Céline Hudelot (2019). “Apprentissage De Relations Floues Pour L’annotation Sémantique Expliquée Avec Peu De Données”. In: *Actes des 17e Rencontres des Jeunes Chercheurs en Intelligence Artificielle 2019*.
- Pino, F., C.L. Fontana, G. Nebbia, C. Carasco, B. Pérot, A. Sardet, J.P. Poli, A.G. Sebert, G. Sannié, A. Iovene, C. Tintori, P. Sibirzynski, K. Grodzicki, L. Swiderski, M. Slegt, R. de Goede, and S. Moretto (Dec. 2022). “Non-intrusive inspection of cargo contain-

- ers using the C-BORD Rapidly Relocatable Tagged Neutron Inspection System”. In: *Journal of Instrumentation* 17.12, T12005.
- Pino, F., C.L. Fontana, G. Nebbia, B. Pedersen, G. Varasano, A. Sardet, C. Carasco, B. Pérot, A.G. Sebert, J.P. Poli, G. Sannié, A. Iovene, C. Tintori, P. Siczynski, K. Grodzicki, L. Swiderski, and S. Moretto (2021). “Detection module of the C-BORD Rapidly Relocatable Tagged Neutron Inspection System (RRTNIS)”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 986, p. 164743. ISSN: 0168-9002.
- Poli, Jean-Philippe (Oct. 2016). “Software Architecture for Expert System”. US Patent App. 15/516,859.
- Poli, Jean-Philippe and Laurence Boudet (2015). “Une architecture moderne de système expert flou pour le traitement des flux d’information”. In: *LFA 2015 - Rencontres francophones sur la Logique Floue et ses Applications*.
- (2016a). “A Modular Fuzzy Expert System Architecture for Data and Event Streams Processing”. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, Cham, pp. 717–728.
- (2016b). “Opérateurs Temporels Flous En Ligne Pour La Comparaison Et La Caractérisation De Signaux”. In: *LFA 2016 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- (2018). “A Fuzzy Expert System Architecture For Data And Event Stream Processing”. In: *Fuzzy Sets and Systems* 343, pp. 20–34.
- Poli, Jean-Philippe, Laurence Boudet, Bruno Espinosa, and Laurence Cornez (2017). “Online Fuzzy Temporal Operators for Complex System Monitoring”. In: *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer, pp. 375–384.
- (2019). “Accesneur Sémantique Aux Données”.
- Poli, Jean-Philippe, Laurence Boudet, and Jean-Marie Le Yaouanc (2018). “Online Spatio-Temporal Fuzzy Relations”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Poli, Jean-Philippe, Laurence Boudet, and David Mercier (2016). “Online Temporal Reasoning for Event and Data Streams Processing”. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 2257–2264.
- Poli, Jean-Philippe, Hiba Hajri, and Laurence Boudet (2021). “Artificial Intelligence for Materials Science and Engineering”. In: *Advanced Manufacturing for Energy and Transportation International School*. EDP Sciences, pp. 311–336.
- Poli, Jean-Philippe and Jean-Paul Laurent (Nov. 2015a). “Assisted Input of Rules into a Knowledge Base”. US Patent App. 14/618,350.
- (2015b). “Interface Tactile Pour La Saisie Guidée De Connaissances”. In: *Proceedings of the 27th Conference on l’Interaction Homme-Machine*. ACM, p. 1.
- (2016). “Touch Interface for Guided Authoring of Expert Systems Rules”. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 1781–1788.
- Poli, Jean-Philippe, Wassila Ouerdane, and Régis Pierrard (2021a). “Génération d’explications textuelles en XAI : le cas de l’annotation sémantique”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- (2021b). “Generation of Textual Explanations in XAI: the Case of Semantic Annotation”. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Quinlan, J. Ross (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Raghu, M., B. Poole, J. Kleinberg, S. Ganguli, and J. Sohl-Dickstein (2017). “On the expressive power of deep neural networks”. In:
- Raj, V. Bhasker, A. T. Nimal, Y. Parmar, M.U. Sharma, and V. Gupta (2012). “Investigations on the origin of mass and elastic loading in the time varying distinct response

- of ZnO SAW ammonia sensor”. In: *Sensors and Actuators B: Chemical* vol. 166.1, p. 573–585.
- Ratle, Alain and Michèle Sebag (2001). “Grammar-guided genetic programming and dimensional consistency: application to non-parametric identification in mechanics”. In: *Applied Soft Computing* 1.1, pp. 105–118.
- Reiter, Ehud and Anja Belz (2009). “An investigation into the validity of some metrics for automatically evaluating natural language generation systems”. In: *Computational Linguistics* 35.4, pp. 529–558.
- Román, Antonio, Abderrahman Fathi, and Julio Santiago (Jan. 2013). “Spatial biases in understanding descriptions of static scenes: The role of reading and writing direction”. In: *Memory & cognition* 41.
- Rotterdam Authority, Port of (2022). *Highlights of the 2018 annual report*. Tech. rep. URL: www.portofrotterdam.com.
- Rousselle, Olivier, Jean-Philippe Poli, and Nadia Ben Abdallah (2023). “Design de plan expérimental optimal : vers une approche basée sur la logique floue”. In: *LFA 2023 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Salleh, Mohd Najib Mohd, Noureen Talpur, and Kashif Hussain (2017). “Adaptive neuro-fuzzy inference system: Overview, strengths, limitations, and solutions”. In: *International Conference on Data Mining and Big Data*. Springer, pp. 527–535.
- Sardet, A., B. Perot, C. Carasco, G. Sannié, S. Moretto, G. Nebbia, C. Fontana, M. Moszyński, P. Sibczyński, K. Grodzicki, L. Świdorski, A. Iovene, and C. Tintori (Oct. 2016). “Design of the rapidly relocatable tagged neutron inspection system of the C-BORD project”. In: *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pp. 1–5.
- Schneider, Patrick and Fatos Xhafa (2022). “Chapter 8 - Machine learning: ML for eHealth systems”. In: *Anomaly Detection and Complex Event Processing over IoT Data Streams*. Ed. by Patrick Schneider and Fatos Xhafa. Academic Press, pp. 149–191.
- Schubert, Erich and Peter J. Rousseeuw (2019). “Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms”. In: *Similarity Search and Applications*. Ed. by Giuseppe Amato, Claudio Gennaro, Vincent Oria, and Miloš Radovanović. Cham: Springer International Publishing, pp. 171–187.
- Shukla, Amit K., Grégory Smits, Olivier Pivert, and Marie-Jeanne Lesot (2020). “Explaining Data Regularities and Anomalies”. In: *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8.
- Smets, Philippe (1997). “Imperfect Information: Imprecision and Uncertainty”. In: *Uncertainty Management in Information Systems: From Needs to Solutions*. Ed. by Amihai Motro and Philippe Smets. Boston, MA: Springer US, pp. 225–254.
- Smits, Grégory, Olivier Pivert, Ronald R. Yager, and Pierre Nerzic (2018). “A soft computing approach to big data summarization”. In: *Fuzzy Sets and Systems* 348, pp. 4–20.
- Sondhi, Parikshit (2009). “Feature construction methods: a survey”. In: *sifaka. cs. uiuc.edu* 69, pp. 70–71.
- Stepin, Iliia, Jose M. Alonso Moral, Alejandro Catala, and Martín Pereira-Fariña (2021). “A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence”. In: *IEEE Access* 9, pp. 11974–12001.
- Subramanian, Girish H., John Nosek, Sankaran P. Raghunathan, and Santosh S. Kanitkar (Jan. 1992). “A Comparison of the Decision Table and Tree”. In: *Commun. ACM* 35.1, pp. 89–94. ISSN: 0001-0782.
- Sun, Ron (2015). “Artificial Intelligence: Connectionist and Symbolic Approaches”. In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Ed. by James D. Wright. Second Edition. Oxford: Elsevier, pp. 35–40.

- Tard, B. (2013). “Etudes des Interactions Gaz-Surfaces Diamant par Gravimétrie sur Résonateur à Onde Acoustique”. PhD thesis. Université Pierre et Marie Curie.
- Thagard, Paul (1989). “Explanatory coherence”. In: *Behavioral and brain sciences* 12.3, pp. 435–467.
- Tsang, Smith, Ben Kao, Kevin Y Yip, Wai-Shing Ho, and Sau Dan Lee (2009). “Decision trees for uncertain data”. In: *IEEE transactions on knowledge and data engineering* 23.1, pp. 64–78.
- Vanegas, M. C. (2011). “Spatial relations and spatial reasoning for the interpretation of Earth observation images using a structural model”. PhD thesis.
- Vanegas, M. C., I. Bloch, and J. Inglada (2016). “Fuzzy constraint satisfaction problem for model-based image interpretation”. In: *Fuzzy Sets and Systems* 286, pp. 1–29.
- Vanegas Orozco, Maria Carolina (Jan. 2011). “Spatial relations and spatial reasoning for the interpretation of Earth observation images using a structural model.” PhD dissertation. Télécom ParisTech.
- Vesky, Jonathon (2008). *Port and maritime security*. New York : Nova Science Publishers.
- Vessey, Iris and Ron Weber (Jan. 1986). “Structured Tools and Conditional Logic: An Empirical Investigation”. In: *Commun. ACM* 29.1, pp. 48–57. ISSN: 0001-0782.
- Vilone, Giulia and Luca Longo (2021). “Notions of explainability and evaluation approaches for explainable artificial intelligence”. In: *Information Fusion* 76, pp. 89–106. ISSN: 1566-2535.
- Vo, Phuc-Nguyen and Marcin Detyniecki (2013). “Towards smooth monotonicity in fuzzy inference system based on gradual generalized modus ponens”. In: *8th conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13)*. Atlantis Press.
- Vo, Phuc-Nguyen, Marcin Detyniecki, and Bernadette Bouchon-Meunier (2013). “Gradual generalized modus ponens”. In: *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 1–7.
- Zadeh, L.A. (1965). “Fuzzy sets”. In: *Information and Control* 8.3, pp. 338–353.
- (1975). “The concept of a linguistic variable and its application to approximate reasoning—I”. In: *Information Sciences* 8.3, pp. 199–249. ISSN: 0020-0255.
- (1983). “A computational approach to fuzzy quantifiers in natural languages”. In: *Computers and Mathematics* 9, pp. 149–184.
- Zadeh, Lotfi A. (1973). “Outline of a New Approach to the Analysis of Complex Systems and Decision Processes”. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.1, pp. 28–44.
- (1996). “Fuzzy logic = computing with words”. In: *IEEE Trans. Fuzzy Syst.* 4, pp. 103–111.
- (1997). “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic”. In: *Fuzzy Sets and Systems* 90.2. Fuzzy Sets: Where Do We Stand? Where Do We Go?, pp. 111–127. ISSN: 0165-0114.
- Zwaan, Rolf A. and Gabriel A. Radvansky (1998). “Situation models in language comprehension and memory.” In: *Psychological Bulletin* 123.2, p. 162.

Appendices

Choose a job you love, and you
will never have to work a day in
your life.

Unknown

Appendix A

Short resume

Appendix B

Teaching activities

I always wanted to teach Computer Science and Artificial Intelligence. Once arrived in Paris, I had the opportunity to teach at Lycée Saint-Louis and then to get in touch with the responsible of the computer science department at Ecole Centrale Paris. All courses are in French except when another language is mentioned.

Teaching at Lycée Saint-Louis

Lycée Saint-Louis is the only school in France with only preparatory classes. From 2003 to 2013, I taught algorithms and Pascal language to the first year students (right after the general Baccalaureate). The goal was to introduce scientific computing and complexity: integrals approximation, zeros finding, polynomials, matrices, etc. It was 64 hours a year, divided into lectures, tutorials and practical work. This first experience allowed me to write a handout for the students and to structure the classes into activities with growing difficulties and synchronized with mathematics classes.

Teaching at Ecole Centrale Paris

In this section, I will present the classes I gave in the former curriculum of Ecole Centrale Paris. Thus, these classes do not exist anymore. I also supervised students during their 3rd year internship.

Algorithms and programming (1st year)

This was the longest class I gave from 2004 to 2015. It has changed according to the responsible of the module, but mainly we taught: python programming, Hoare's logic, testing, complexity, sorting, trees, and graphs. During my PhD, I helped with tutorials and practical work. Then, I also gave lectures and actively participated to the writing of exams and tutorials.

Introduction to Artificial Intelligence (3rd year)

From 2011 to 2020, this class was an overview of AI, from knowledge representation and problem solving to machine learning. Each class of 3h was split into a lecture of more or less 45 minutes, 45 minutes of tutorials and 1h30 of practical work. I started with lectures of 90 minutes but I realized the students did not focus more than 45 minutes. Students appreciated such a distribution. I also invented a board game called "Vampires vs Werewolves" that can be played by alpha-beta based methods if the combinatory is correctly tackled. At the end of the class (27 hours per year), a tournament was organized with all the students' AIs competing. The competition was a relaxed moment to share

with the students. I also tested the students with a 3 hours exam whose exercises were themed regarding my personal life or TV shows.

Object-Oriented Advanced Programming (3rd year)

I designed this new type of programming class whose goal was to use the latest techniques in programming. Since the levels of the students were heterogeneous, the typical class with lectures was not possible: the most advanced students would have been bored. Moreover, the future careers of these students were also heterogeneous. I decided to let the students choose among different programming languages (Python, Java, C#). I set up a flipped classroom with three phases:

1. Skill development: I gave online resources to learn the syntax of each language. The students had to work at home.
2. Tutorials: each student had 7.5 hours (3 hours + 3 × 1.5 hours) in class to complete tutorials on object-oriented modeling, multithreading, sockets and webservices.
3. Projects: the remainder of the hours were kept for a project. The students were split into groups of 3 students. I gave a list of projects and let the students suggest their own if they wanted.

Partners of the school were helping with this module, thus offering a high supervision rate. Each project was evaluated regarding its originality, difficulty and quality of code. The students were also evaluated with an online programming exam (with questions such as in job interviews). This class was the longest in terms of hours: 36 hours per year.

Software development (3rd year)

After the previous class, we focused in this one on the quality of code (documentation, code review, git, SCRUM method, etc). I animated this class with Kevin Nguyen, from *The coding machine*. Each theme was developed during a short lecture of 45 min. The audience was split into groups of 6 to 8 students to develop a service. I also gave a list of projects and the students were invited to propose their own projects. It was 24 hours per year of lectures and practical work. Each group was evaluated during a defense.

Projects supervision (1st and 2nd years)

I supervised projects in computer science. For the 1st year students, the goal was to develop a prototype for a company. It could be developing algorithms, like computer vision, or developing a smartphone application. For the 2nd year, it was innovation projects, with more hours.

Teaching at CentraleSupélec

In this section, I will describe the classes to which I participate in the new curriculum. I also supervised students during their 3rd year internship.

Co-supervision of AI projects (1st and 2nd years)

The school asks each student to participate to projects. I co-supervise all the projects with my colleague Wassila Ouerdane. The goal is the same as in the former curriculum: students have to implement an AI solution for a company. It can be deep learning methods as well as recommender systems.

Foundation of Artificial Intelligence (MSc in AI)

From 2018, I teach this class in the Master of Science in Artificial Intelligence, with international students (the class is thus in English). It is an introduction to AI in general, and a focus on symbolic AI. The students are evaluated on a challenge with a modified version of Vampires vs Werewolves.

Computational approach to games (1st year)

I am not responsible of this module but I teach the AI part (solving, constraint programming, alpha-beta, Monte-Carlo tree search). I give the lectures (1.5 hour) and the tutorials (3 hours).

AI challenge (3rd year)

This short module consists in applying AI techniques on a problem that is given by a partner of the school. The audience is split into groups of 5 students.

Explainable Artificial Intelligence (3rd year)

I give lectures and tutorials (12 hours per year) in this module that matches my research interests. It is about transparent models and post-hoc explanation.

Teaching at INSTN

INSTN is an institute attached to CEA. I teach an introduction to Artificial Intelligence for material science in the master of material for energy and transportation (Master M2 MET). The goal is to give an overview of the main methods that are used in material science, mainly data driven and machine learning methods. It represents 3 hours per year of lecture.

Summer schools

I participate to the AMETIS summer school about AI for material science and to the summer school in Artificial Intelligence of CentraleSupélec where I introduce AI, symbolism and connectionism.

Trainings

I also animate trainings about Artificial Intelligence, of different durations. For instance, during the Digital Tech Year, it is one week to introduce all AI. For companies, as Thales and Naval Group, it is one day for an introduction to symbolic AI.

Apprentices and interns

I supervised two apprentices as an academic supervisor at CentraleSupélec: Erwan Mahé, Owain Biddulph.

I also supervised several developer apprentices from Polytech' Paris Saclay: Stéphane Barat, Jérémy Babouche, Florent Brouca, Robin Delgado, Bastien Guillon, Mikaël Gendreau, Sébastien Klasa, Lucas Payet, Kevin Gallus, Julie Dornat and Ali Mahmoud.

I supervised several interns during their second year in an engineering school: Tristan Galliez, Etienne Pétreil, Théo Rubenach, Julien Plouvier, Sylvain Jankowiak, Martin Everaert.

Appendix C

Collective responsibilities

Since my doctorate, I have participated as a researcher to many activities.

PhD jury member

I had the opportunity to be a member of the jury of the two following PhD defenses:

- Tanguy Giuffrida: *Fuzzy4U : un système d'adaptation des Interfaces Homme-Machine en logique floue pour l'accessibilité*, Université Grenoble Alpes, in December 2020.
- Elie Daher: *A participatory method and toolset for data-driven optimization of design solutions in parametric modeling systems*, KU Leuven, in December 2020.

Conference committees

I had the opportunity to be a member of the technical committees of EuroSensors 2017, when it has been organized in Paris, and since 2021, of the french conference on fuzzy logic, LFA (Rencontres francophones sur la logique floue et ses applications).

Reviews

I also review papers from different conferences and journals:

- In 2013: ANR projects about security.
- In 2017: EuroSensors.
- In 2023: Innovation in Humanitarian Habitat (Croix-Rouge luxembourgeoise), Qeios, AI Communications.
- Since 2021: FuzzIEEE.
- Since 2021: LFA (Rencontres francophones sur la logique floue et ses applications).
- Since 2018: IJCAI, AAAI.
- Since 2018: different Elsevier journals, as Fuzzy Sets and Systems, Expert Systems with Applications, Computer Methods and Programs in Biomedicine, Machine Learning with Applications.
- Since 2007: Multimedia Tools and Applications.

Other

In November 2019, I have been invited to talk about AI and transportation at ZeBox, CMA-CGM incubator in Marseilles. In June 2022, I have been invited by the Research Center of the French National Police Department to talk about the pros and the cons of AI for police.

I have also been invited to be part of the improvement committee of ENSIL-ENSCI engineering school in 2022. This year, the topic was Artificial Intelligence and the committee aims at improving the curriculum of the school to fit the needs of industry. In March 2023, I have been invited by ENSIL-ENSCI to give a lecture in AI for material science.

In November 2022, I have been invited to present XAI to a Schlumberger lab. In March 2023, I presented within the French GDR-ISIS my work at the interface between deep learning and symbolic AI for computer vision.

I also follow some students of CentraleSupélec during their final internship (3-4 students per year).

Appendix D

Supervision of young researchers

M2 Interns

Jean-Paul Laurent

Topic: Rule authoring on touch screens
Degree: M2 Ergonomics, University of Paris Sud
Dates: April - September 2013
Publications: Poli and Laurent, 2015a,b, 2016

Rabah Abdul Khalek

Topic: Neural networks for the recognition of DVCS events
Degree: M2 Experimental Physics, Université Paris Saclay
Dates: May - September 2017
Co-supervisors: Maxime Defurne, Franck Sabatié

Ismail Baaj

Topic: Explaining fuzzy inference systems decisions
Degree: M2 Distributed Artificial Intelligence, Université Paris Descartes
Dates: April - September 2018
Publication: Baaj and Poli, 2019

Thomas Lamson

Topic: Knowledge discovery and active learning
Degree: Master of Science in Artificial Intelligence, CentraleSupélec
Dates: April - September 2020

Killian Susini

Topic: Study of graduality in fuzzy logic
Degree: Master of Science in Artificial Intelligence, CentraleSupélec
Dates: April - September 2022

Antonin Arsac

Topic: Fuzzy rule learning for the prediction of time series
Degree: M2 MApI3, Université Paul Sabatier
Dates: April - September 2022
Co-supervisor: Aurore Lomet
Publication: Arsac, Lomet, and Poli, 2023

Postdoctoral fellows and other non-permanent researchers

Jean-Marie le Yaouanc

Topic: Activity recognition with fuzzy spatio-temporal relations
Dates: 2011 - 2012
Funding: Partnership with Egidium
Publications: Le Yaouanc and Poli, 2012; Poli, Boudet, and Le Yaouanc, 2018
Current position: Patent engineer at Cabinet Beau de Loménie

Arnaud Grivet Sébert

Topic: Fuzzy rule learning from uncertain data
Dates: 2017 - 2018
Funding: European project *C-BORD*
Publications: Grivet Sébert and Poli, 2017, 2018a,b
Current position: PhD candidate at CEA List

Marouen Baalouch

Topic: Neural networks for the recognition of DVCS events
Dates: 2018
Co-supervisor: Maxime Defurne
Funding: CEA Cross-Cutting Program on simulation *Alcryphe*
Publication: Baalouch et al., 2019
Current position: Lead data scientist at Capgemini

Hiba Hajri

Topic: Learning fuzzy gradual rules for material science
Dates: 2019-2021
Co-supervisor: Laurence Boudet
Funding: CEA Cross-Cutting Program on materials *ProvidIA*
Publications: Hajri, Poli, and Boudet, 2021b,a; Poli, Hajri, and Boudet, 2021
Current position: R& D consultant at Business & Decision

Clément Iphar

Topic: Fuzzy spatial relations for natural crisis management
Dates: 2021
Co-supervisor: Laurence Boudet
Funding: European project *Respondrone*
Publications: Iphar, Boudet, and Poli, 2021b,a, 2023b
Current position: Post-doctorant at Université de Bretagne Occidentale

Lucie Kunitomo Jacquin

Topic: Extraction of fuzzy causal relations from few data
Dates: 2021-2023
Co-supervisor: Aurore Lomet
Funding: CEA Cross-Cutting Program on materials *C3PO*
Publications: Jacquin, Lomet, and Poli, 2021; Kunitomo-Jacquin, Lomet, and Poli, 2022a,b
Current position: Post-doctorant at the National Institute of Advanced Industrial Science and Technology (Japan)

Nadia Ben Abdallah

Topic: Extraction of fuzzy causal relations from few data
Dates: 2021 (hired as permanent researcher in ExpressIF team)
Funding: CEA Cross-Cutting Program on materials *dIAmant*
Current position: Permanent researcher at CEA List

Olivier Rousselle

Topic: Optimal design of experiments
Dates: 2023-2024
Co-supervisor: Nadia Ben Abdallah
Funding: CEA Cross-Cutting Program on materials *dIAmant*
Publications: Rousselle, Poli, and Ben Abdallah, 2023

PhD candidates

Olivier Hotel

Title: Algorithms, Methods and Models for the Application of Surface Acoustic Waves Sensors to the Recognition of Chemical Compounds Signatures
Defense: on December 11th, 2017
Director: Samuel Saada (CEA)
Supervision rate: 80%
Publications: Friedmann, Poli, et al., 2020; Hotel, Poli, Mer-Calfati, et al., 2017a,b, 2018; Hotel, Poli, and Saada, 2017

Recently, gas sensor arrays have found numerous applications in areas such as the food, the environment, the medicine and the defense industries. Among the existing technologies, the surface acoustic wave technology is one of the most promising and has been the subject of abundant research. The work described in this manuscript concerns the development of algorithms allowing the recognition of chemical compounds and the estimation of their concentration. This study describes a method for estimating the parameters of transduction phenomena. Their interest is demonstrated experimentally in applications consisting in identifying toxic chemical compounds, counterfeit coffee capsules and in detecting the presence of DMMP and 4-NT in the presence of interfering compounds.

Olivier Hotel is currently working as postdoctoral fellow at Grenoble INP.

Régis Pierrard

Title: Explainable Classification and Annotation through Relation Learning and Reasoning
Defense: on September 15th, 2020
Director: Céline Hudelot (CentraleSupélec)
Supervision rate: 80%
Publications: Pierrard, Cabaret, et al., 2020; Pierrard, Poli, and Hudelot, 2018a,b, 2019a, 2021; Pierrard, Poli, and Hudelot, 2019b; Poli, Ouerdane, and Pierrard, 2021b,a

With the recent successes of deep learning and the growing interactions between humans and AIs, explainability issues have risen. Indeed, it is difficult to understand the behaviour of deep neural networks and thus such opaque models are not suited for high-stake

applications. In this thesis, we propose an approach for performing classification or annotation and providing explanations. It is based on a transparent model, whose reasoning is clear, and on interpretable fuzzy relations that enable to express the vagueness of natural language. Instead of learning on training instances that are annotated with relations, we propose to rely on a set of relations that was set beforehand. We present two heuristics that make the process of evaluating relations faster. Then, the most relevant relations can be extracted using a new fuzzy frequent itemset mining algorithm. These relations enable to build rules, for classification, and constraints, for annotation. Since the strengths of our approach are the transparency of the model and the interpretability of the relations, an explanation in natural language can be generated. We present experiments on images and time series that show the genericity of the approach. In particular, the application to explainable organ annotation was received positively by a set of participants that judges the explanations consistent and convincing.

Régis Pierrard is currently researcher at Hugging face.

Noélie Cherrier

Title: Interpretable Machine Learning for CLAS12 Data Analysis
 Defense: on March 1st, 2021
 Director: Franck Sabatié (CEA)
 Co-supervisor: Maxime Defurne (CEA)
 Supervision rate: 40%
 Publications: Baalouch et al., 2019; Cherrier, Defurne, et al., 2019; Cherrier, Mayo, et al., 2020; Cherrier, Poli, Defurne, and Sabatié, 2019, 2020

Artificial intelligence is used massively in numerous applications, especially since the rise of deep learning techniques. However, some of these applications require a careful study and validation of the inducted model functioning. Considering experimental physics, the performances of the models on real data must be known and controlled, and their functioning explained to enable a validation via peer review. In the particular case of the CLAS12 experiment at Jefferson Laboratory, an electron beam is sent onto a proton target to probe its inner structure. To access certain structure functions of the proton, a subset of the collected data must be selected corresponding to an exclusive interaction: deeply virtual Compton scattering. This thesis focuses on this event selection. To improve the classical physics analysis, an approach exploiting intrinsically interpretable machine learning models, also called transparent models, is proposed. In this way, the functioning of the model is understood more easily and the selection errors are minimized and controlled.

Noélie Cherrier is currently researcher at CITiO.

Ismail Baaj

Title: Explainability of possibilistic and fuzzy rule-based systems
 Defense: January 27th, 2022
 Director: Nicolas Maudet (Sorbone Université)
 Co-supervisor: Wassila Ouerdane (CentraleSupélec)
 Supervision rate: 40%
 Publications: Baaj, Poli, Ouerdane, and Maudet, 2021b,a,c; Baaj and Poli, 2019; Baaj, Poli, and Ouerdane, 2019

The explanatory capability of AI systems has become a user requirement, especially in human-risk environments such as autonomous vehicles or medicine. This requirement is in line with the recent resurgence of interest for eXplainable Artificial Intelligence (abbreviated as XAI), a research field that aims to develop AI systems that are able to explain their results in a way that is comprehensible to humans. We introduce methods for justifying the inference results of possibilistic and fuzzy rule-based systems. Our methods lead to form two kinds of explanations of an inference result of these systems: 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). Finally, we propose a graphical representation of an explanation of an inference result of a possibilistic or fuzzy rule-based system in terms of conceptual graphs. For an inference result, we represent its justification, its unexpectedness and a combination of its justification and its unexpectedness.

Ismail Baaj is currently in a postdoctoral fellowship at CRIL, University of Artois.

Antonin Arsac

Title:	Causal fuzzy rule-based prediction of time series based
Start:	November, 2022
Director:	Jean-Philippe Poli (CEA List)
Supervisor:	Aurore Lomet (CEA List)
Supervision rate:	20%
Publication:	Arsac, Lomet, and Poli, 2023

This thesis aims at developing an innovative solution for learning a rule base for time series prediction. To do so, it is first necessary to have an automatic method to deduce the causal relationships between time series. In a second step, it will be necessary to apply or adapt or develop vocabulary extraction methods for the definition of rules from established causal relationships. New operators specific to time series can be defined. The different types of fuzzy rules can be investigated.

Appendix E

Research projects and fundings

Due to the nature itself of CEA Tech, we always need collaborations to carry out our research. The most important partnerships are industrial ones, because it matches totally with the missions of CEA Tech.

Industrial partnerships

In this section, I focus on the industrial partnerships. The research is here transferred to the partners thanks to the ExpressIF[®] software.

Table ?? summarizes the industrial partnerships since 2012. I indicate the topics of the projects since I cannot give more details about those projects. I indicate the number of persons of ExpressIF team that are involved in the project, including me. I also indicate if I was the project leader and if I did the prospecting that led to the project. In all those projects, I had a scientific and technical contribution.

European projects

European projects are important to benefit from a longer time to work on a given topic. It is also the occasion to mix different sciences, different know-how, different approaches and to discover different cultures. At the end of the projects, demonstrations are needed so European projects are also a great opportunities for implementations. Table E.1 shows the European projects I have been implied in. For each of them, I give the starting year, the application domain and I indicate whether I participated to the project proposal, and whether I was project manager at CEA's level, work package leader, or task leader. I was implied technically or scientifically in all of them. All those projects lastes three years, and sometimes, due to the pandemic, 3.5 years.

National projects

I also participated to national funded projects. Mobisic was the one I was hired for during my postdoctoral fellowship. The three of them allowed building ExpressIF[®] and so my scientific activities.

Table E.2 presents the three national projects. I remind the starting year, the application domain, and whether I was the project leader at CEA's level.

Internal fundings

The CEA has many mechanisms to encourage innovation. In particular, the Cross-Cutting Program allows funding PhDs, postdoctoral fellowships, etc. It is a two-step process: a

Year	Acronym	Application	Proposal	Project manager	Work package leader	Task leader
2011	Secured	Security				
2012	Scintilla	Security				
2013	Safewater	Security				
2015	BIM4VET	Building		✓		
2015	CBORD	Security		✓		✓
2017	BIMEET	Building		✓		✓
2018	NARSIS	Crisis management	✓	✓		✓
2018	Terrific	Security				
2019	Spadassin	Security				
2019	DeepHealth	Health	✓	✓		✓
2019	Respondrone	Crisis management	✓	✓	✓	✓
2019	MICADO	Dismantling	✓			✓
2020	Entrance	Security	✓	✓		✓
2020	DetecTool	Health				
2022	Including	Security				

Table E.1: European projects.

Year	Acronym	Application domain	Project Manager
2007	MobiSIC	Security	✓
2008	Descartes	Crisis management	
2010	Edens	Energetic transition	✓

Table E.2: National projects.

first selection based on a summary, budget, motivation and state-of-the-art, and the last selection based on a more complete proposal. Table E.3 gives some information about the 4 grants I obtained since 2017.

Years	Acronym	Application domain	Proposal	Project manager
2017 – 2020	Alcryphe	Physics	✓	✓
2019 – 2021	ProvidIA	Materials	✓	✓
2020 – 2024	DIAMANT	Materials	✓	✓
2020 – 2023	C3PO	Materials	✓	✓
2023 – 2025	ExIL	Materials	✓	✓
2023 – 2025	CHIPS	Materials		
2023 – 2026	PopCorn	Materials		

Table E.3: CEA projects.

Some CEA divisions can also directly fund activities of other ones. One of my post-doctoral projects was funded by another division of CEA (CLARISSE project). I worked two years on CLARISSE that was about automatically locating the epicenter of a seism. DGA (the French Defence Procurement Agency) funds the Essaim project. Essaim aims at building a new generation of chemical sensor based on SAW technology. The project started in 2015 and has been renewed every year since.

Appendix F

Complete list of publications

You will find in this section all the publications I co-authored. Table F.1 gives an overview per categories. Note that contrary to computer science, there is no such classification as A*, A, etc., in physics.

Categories	Total	A*	A	B
International journals	7	2	2	2
National journals	1			
International conferences with proceedings	32		6	12
International conferences without proceedings	3			
French conferences with proceedings	19			
French conferences without proceedings	1			
International workshops with proceedings	5			
French workshops without proceedings	1			
Book chapters	1			
PhD	1			
Patents	5			
Total	76	2	8	14

Table F.1: Publications per categories.

International journals

- Pino, F., C.L. Fontana, G. Nebbia, C. Carasco, B. Pérot, A. Sardet, J.P. Poli, A.G. Sebert, G. Sannié, A. Iovene, C. Tintori, P. Sibczynski, K. Grodzicki, L. Swiderski, M. Slegt, R. de Goede, and S. Moretto (Dec. 2022). “Non-intrusive inspection of cargo containers using the C-BORD Rapidly Relocatable Tagged Neutron Inspection System”. In: *Journal of Instrumentation* 17.12, T12005.
- Pierrard, Régis, Jean-Philippe Poli, and Céline Hudelot (2021). “Spatial relation learning for explainable image classification and annotation in critical applications”. In: *Artificial Intelligence* 292, p. 103434. ISSN: 0004-3702.
- Pino, F., C.L. Fontana, G. Nebbia, B. Pedersen, G. Varasano, A. Sardet, C. Carasco, B. Pérot, A.G. Sebert, J.P. Poli, G. Sannié, A. Iovene, C. Tintori, P. Sibczynski, K. Grodzicki, L. Swiderski, and S. Moretto (2021). “Detection module of the C-BORD Rapidly Relocatable Tagged Neutron Inspection System (RRTNIS)”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 986, p. 164743. ISSN: 0168-9002.

- Hotel, Olivier, Jean-Philippe Poli, Christine Mer-Calfati, Emmanuel Scorsone, and Samuel Saada (2018). “A Review Of Algorithms For Saw Sensors E-nose Based Volatile Compound Identification”. In: *Sensors and Actuators B: Chemical* 255, pp. 2472–2482.
- Poli, Jean-Philippe and Laurence Boudet (2018). “A Fuzzy Expert System Architecture For Data And Event Stream Processing”. In: *Fuzzy Sets and Systems* 343, pp. 20–34.
- Hotel, Olivier, Jean-Philippe Poli, Christine Mer-Calfati, Emmanuel Scorsone, and Samuel Saada (2017b). “Saw Sensor’s Frequency Shift Characterization for Odor Recognition and Concentration Estimation”. In: *IEEE Sensors Journal* 17.21, pp. 7011–7018.
- Poli, Jean-Philippe (2008). “An Automatic Television Stream Structuring System for Television Archives Holders”. In: *Multimedia systems* 14.5, pp. 255–275.

French journals

- Giuffrida, Tanguy, Eric Céret, Sophie Dupuy-Chessa, and Jean-Philippe Poli (2019). “Fuzzy4u: un Moteur d’Adaptation en Logique Floue pour l’Accessibilité ses Interfaces Utilisateurs”. In: *Journal d’Interaction Personne-Système* 8.1, pp. 27–59.

International conferences with proceedings

- Iphar, Clément, Laurence Boudet, and Jean-Philippe Poli (2023a). “Topography-based Fuzzy Assessment of Burning Area in Wildfire Spread Simulation”. In: *Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)*.
- Kunitomo-Jacquín, Lucie, Aurore Lomet, and Jean-Philippe Poli (2022a). “Causal discovery for fuzzy rule learning”. In: *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Baaj, Ismaïl, Jean-Philippe Poli, Wassila Ouerdane, and Nicolas Maudet (2021a). “Min-max inference for Possibilistic Rule-Based System”. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- (2021c). “Representation of Explanations of Possibilistic Inference Decisions”. In: *Proceedings of ECSQARU*.
- Hajri, Hiba, Jean-Philippe Poli, and Laurence Boudet (2021b). “Towards Monotonous Functions Approximation from Few Data With Gradual Generalized Modus Ponens: Application to Materials Science”. In: *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 796–800.
- Iphar, Clément, Laurence Boudet, and Jean-Philippe Poli (2021a). “Topography-based Fuzzy Assessment of Runoff Area with 3D Spatial Relations”. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Poli, Jean-Philippe, Wassila Ouerdane, and Régis Pierrard (2021a). “Generation of Textual Explanations in XAI: the Case of Semantic Annotation”. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Boudet, Laurence, Jean-Philippe Poli, Louis-Pierre Bergé, and Michel Rodriguez (2020). “Situational Assessment of Wildfires: a Fuzzy Spatial Approach”. In: *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 1180–1185.
- Cherrier, Noëlie, Michael Mayo, Jean-Philippe Poli, Maxime Defurne, and Franck Sabatié (2020). “Interpretable Machine Learning with Bitonic Generalized Additive Models and Automatic Feature Construction”. In: *Discovery Science*. Ed. by Annalisa Appice, Grigorios Tsoumakas, Yannis Manolopoulos, and Stan Matwin. Cham: Springer International Publishing, pp. 386–402.
- Cherrier, Noëlie, Jean-Philippe Poli, Maxime Defurne, and Franck Sabatié (2020). “Embedded Feature Construction in Fuzzy Decision Tree Induction for High Energy Physics

- Classification”. In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 615–622.
- Friedmann, Edwin, Jean-Philippe Poli, Olivier Hotel, and Christine Mer-Calfati (2020). “Fuzzy Classifiers for Chemical Compound Recognition from SAW Sensors Signals”. In: *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 917–922.
- Baaï, Ismaïl and Jean-Philippe Poli (2019). “Natural Language Generation of Explanations of Fuzzy Inference Decisions”. In: *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Cherrier, Noëlie, Jean-Philippe Poli, Maxime Defurne, and Franck Sabatié (2019). “Consistent Feature Construction with Constrained Genetic Programming for Experimental Physics”. In: *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, pp. 1650–1658.
- Giuffrida, Tanguy, Sophie Dupuy-Chessa, Jean-Philippe Poli, and Éric Céret (2019). “Fuzzy4u: A Fuzzy Logic System for User Interfaces Adaptation”. In: *2019 13th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, pp. 1–12.
- Guerriero, Annie, Sylvain Kubicki, V Maquil, N Mack, Yacine Rezgui, H Li, S Lamb, A Bradley, and J-P Poli (2019). “BIM4VET, Towards BIM Training Recommendation for AEC Professionals”. In: *Advances in Informatics and Computing in Civil and Construction Engineering*. Springer, pp. 833–840.
- Boudet, Laurence, Jean-Philippe Poli, Alicia Bel, François Castillon, Frédéric Gaigne, and Olivier Casula (2018). “Design of a Decision Support System for Buried Pipeline Corrosion Assessment”. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, pp. 74–85.
- Grivet Sébert, Arnaud and Jean-Philippe Poli (2018a). “Fuzzy Rule Learning for Material Classification from Imprecise Data”. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, pp. 62–73.
- (2018b). “Material Classification from Imprecise Chemical Composition: Probabilistic vs Possibilistic Approach”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Pierrard, Régis, Jean-Philippe Poli, and Céline Hudelot (2018a). “A Fuzzy Close Algorithm for Mining Fuzzy Association Rules”. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, pp. 88–99.
- (2018b). “Learning Fuzzy Relations and Properties for Explainable Artificial Intelligence”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 1–8.
- Poli, Jean-Philippe, Laurence Boudet, and Jean-Marie Le Yaouanc (2018). “Online Spatio-Temporal Fuzzy Relations”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Hotel, Olivier, Jean-Philippe Poli, Christine Mer-Calfati, Emmanuel Scorsone, and Samuel Saada (2017a). “Estimation of the Parameters of Saw Sensor’s Frequency Shift: Application to Odour Recognition and Concentration Evaluation”. In: *2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*. IEEE, pp. 1–3.
- (2017c). “Estimation of the Number of Volatile Compounds in Simple Mixtures”. In: *Proceedings of EuroSensors 2017*. Vol. 1. 4, p. 623.
- Poli, Jean-Philippe, Laurence Boudet, Bruno Espinosa, and Laurence Cornez (2017). “Online Fuzzy Temporal Operators for Complex System Monitoring”. In: *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Springer, pp. 375–384.

- Bel, Alicia, Laurence Boudet, Jean-Philippe Poli, François Castillon, Frédéric Graigne, and Olivier Casula (2016). “A Fuzzy Expert System for Buried Pipeline Corrosion Assessment Based on Dcvg Measurements”. In: *Eurocorr2016*.
- Poli, Jean-Philippe and Laurence Boudet (2016a). “A Modular Fuzzy Expert System Architecture for Data and Event Streams Processing”. In: *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, Cham, pp. 717–728.
- Poli, Jean-Philippe, Laurence Boudet, and David Mercier (2016). “Online Temporal Reasoning for Event and Data Streams Processing”. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 2257–2264.
- Poli, Jean-Philippe and Jean-Paul Laurent (2016). “Touch Interface for Guided Authoring of Expert Systems Rules”. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, pp. 1781–1788.
- Poli, Jean-Philippe and Jean Carrive (2007). “Modeling Television Schedules for Television Stream Structuring”. In: *International Conference on Multimedia Modeling*. Springer, Berlin, Heidelberg, pp. 680–689.
- (2006a). “Television Stream Structuring with Program Guides”. In: *Eighth IEEE International Symposium on Multimedia (ISM’06)*. IEEE, pp. 329–334.
- (2006b). “Improving Program Guides for Reducing Tv Stream Structuring Problem to a Simple Alignment Problem”. In: *2006 International Conference on Computational Intelligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA’06)*. IEEE, pp. 31–31.
- Poli, J-P (2005). “Predicting Program Guides for Video Structuring”. In: *17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’05)*. IEEE, 5–pp.

International conferences without proceedings

- Magne, Sylvain, Wilfrid Husson, Grigaut-Desbrosses Hans Payet Lucas, Guillaume Sutra, Stéphane Dogny, Pierre-Guy Allineï, Marion Ledieu, Olivier Guéton, and Bernard Leibovici (2021). “1D OSL/FO Dosimeter Array for Remote Radiological Investigations in Hard-to-access Zones”. In: *The 7th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications*.
- Sari, Adrien, Frédérick Carrel, Amélie Grabowski, Frédéric Lainé, Bruno Espinosa, Jean-Philippe Poli, Pawel Sibczyński, Ian Della-Rocca, Mark Foster, Asénath Etilé, Olivier Roig, Serge Maitrejean, Sébastien Rogerat, Thibaut Berthelie, Estelle Gasser, Micha Slegt, René de Goede, Joris Groeneveld, Hans de Wilde, and Marcel Heerschop (2019). “Deployment of the First Photofission Measurement System Dedicated to SNM Detection in Europe: Outcomes and Future Prospects”. In: *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*.
- Boudergui, K, F Carrel, T Domenech, N Guenard, J-P Poli, A Ravet, V Schoepff, and R Woo (2011). “Development of a Drone Equipped with Optimized Sensors for Nuclear and Radiological Risk Characterization”. In: *2011 2nd International Conference on Advancements in Nuclear Instrumentation, Measurement Methods and Their Applications*.

French conferences with proceedings

- Iphar, Clément, Laurence Boudet, and Jean-Philippe Poli (2023b). “Evaluation floue d’une zone de feu basée sur la topographie pour la simulation de la propagation de feux de forêt”. In: *LFA 2023 - Rencontres Francophones sur la Logique Floue et ses Applications*.

- Rousselle, Olivier, Jean-Philippe Poli, and Nadia Ben Abdallah (2023). “Design de plan expérimental optimal : vers une approche basée sur la logique floue”. In: *LFA 2023 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Kunitomo-Jacquín, Lucie, Aurore Lomet, and Jean-Philippe Poli (2022b). “Exploitation de la causalité pour l’apprentissage de règles floues”. In: *LFA 2022 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Baaï, Ismaïl, Jean-Philippe Poli, Wassila Ouerdane, and Nicolas Maudet (2021b). “Inférence min-max pour un système à base de règles possibilistes”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Hajri, Hiba, Jean-Philippe Poli, and Laurence Boudet (2021a). “Vers l’approximation de fonctions monotones avec le modus ponens généralisé graduel à partir de peu de données : Application à la Science des Matériaux”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Iphar, Clément, Laurence Boudet, and Jean-Philippe Poli (2021b). “Evaluation d’une zone d’écoulement floue basée sur la topographie à l’aide de prédicats spatiaux 3D”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Jacquín, Lucie, Aurore Lomet, and Jean-Philippe Poli (2021). “Discussion Sur La Causalité Imparfaite Et Son Application Aux Sciences Expérimentales”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Poli, Jean-Philippe, Wassila Ouerdane, and Régis Pierrard (2021b). “Génération d’explications textuelles en XAI : le cas de l’annotation sémantique”. In: *LFA 2021 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Boudet, Laurence, Jean-Philippe Poli, Pierre-Louis Bergé, and Michel Rodriguez (2019). “Règles spatiales floues et SIG pour l’évaluation d’un risque : le cas des feux de forêt”. In: *LFA 2019 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Friedmann, Edwin and Jean-Philippe Poli (2019). “Système à base de règles floues pour la reconnaissance de composés chimiques”. In: *LFA 2019 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- Pierrard, Régis, Jean-Philippe Poli, and Celine Hudelot (2019b). “Apprentissage De Relations Floues Pour L’annotation Sémantique Expliquée Avec Peu De Données”. In: *Actes des 17e Rencontres des Jeunes Chercheurs en Intelligence Artificielle 2019*.
- Giuffrida, Tanguy, Sophie Dupuy-Chessa, Jean-Philippe Poli, and Eric Ceret (2018). “Fuzzy4U: un système en logique floue pour l’adaptation des interfaces utilisateur”. In: *30ème conférence francophone sur l’Interaction Homme-Machine IHM-2018*.
- Grivet Sébert, Arnaud and Jean-Philippe Poli (2017). “Vers la Classification de Matériaux à Partir de Compositions Chimiques Incertaines”. In: *LFA 2017 - Rencontres francophones sur la Logique Floue et ses Applications*.
- Poli, Jean-Philippe and Laurence Boudet (2016b). “Opérateurs Temporels Flous En Ligne Pour La Comparaison Et La Caractérisation De Signaux”. In: *LFA 2016 - Rencontres Francophones sur la Logique Floue et ses Applications*.
- (2015). “Une architecture moderne de système expert flou pour le traitement des flux d’information”. In: *LFA 2015 - Rencontres francophones sur la Logique Floue et ses Applications*.
- Poli, Jean-Philippe and Jean-Paul Laurent (2015b). “Interface Tactile Pour La Saisie Guidée De Connaissances”. In: *Proceedings of the 27th Conference on l’Interaction Homme-Machine*. ACM, p. 1.
- Poli, Jean-Philippe, David Mercier, Anthony Larue, Carole Maillard, and Jocelyn Guibert (2009). “Génération Rapide de Scénarios Géophysiques par Satisfaction de Contraintes pour la Localisation des Séismes”. In: *Cinquièmes Journées Francophones de Programmation par Contraintes, Orléans, juin 2009*, 10–pages.
- Poli, Jean-Philippe and Jean Carrive (2006c). “Prédiction de Séries Temporelles: Application à la Structuration des Flux Audiovisuels”. In:

Poli, Jean-Philippe and Jean Carrive (2005a). “Proposition d’une Architecture pour un Système de Structuration de Flux Audiovisuels”. In: *Actes des Journées CORESA*. CORESA.

French conferences without proceedings

Hotel, Olivier, Jean-Philippe Poli, and Samuel Saada (2017). “Apports De L’optimisation Par Essaims Particulaires Pour L’identification De Composés Chimiques”. In: *18ème Conférence ROADEF de la Société Française de Recherche Opérationnelle et Aide à la Décision*.

International workshops with proceedings

Pierrard, Régis, Laurent Cabaret, Jean-Philippe Poli, and Céline Hudelot (2020). “SIMD-Based Exact Parallel Fuzzy Dilation Operator for Fast Computing of Fuzzy Spatial Relations”. In: *Proceedings of the 2020 Sixth Workshop on Programming Models for SIMD/Vector Processing*. WPMVP’20. Association for Computing Machinery.

Pierrard, Régis, Jean-Philippe Poli, and Céline Hudelot (2019a). “A New Approach for Explainable Multiple Organ Annotation with Few Data”. In: *IJCAI 2019 Workshop on Explainable AI*, pp. 101–107.

Le Yaouanc, Jean-Marie and Jean-Philippe Poli (2012). “A Fuzzy Spatio-temporal-based Approach for Activity Recognition”. In: *International Conference on Conceptual Modeling*. Springer, pp. 314–323.

Poli, Jean-Philippe, Julien Piquier, Jean Carrive, and Jérémy Philippeau (2007). “Fast Hierarchical Multimodal Structuring of Time Slots”. In: *2007 International Workshop on Content-Based Multimedia Indexing*. IEEE, pp. 77–84.

Troncy, Raphaël, Jean Carrive, Steffen Lalande, and Jean-Philippe Poli (2004). “A Motivating Scenario for Designing an Extensible Audio-Visual Description”. In: *In The International Workshop on Multidisciplinary Image, Video, and Audio Retrieval and Mining (CoRIMedia)*.

French workshops without proceedings

Arsac, Antonin, Aurore Lomet, and Jean-Philippe Poli (2023). *Causal discovery for time series with constraint-based model and PMIME measure*. arXiv: 2305.19695 [stat.ME].

Book chapters

Poli, Jean-Philippe, Hiba Hajri, and Laurence Boudet (2021). “Artificial Intelligence for Materials Science and Engineering”. In: *Advanced Manufacturing for Energy and Transportation International School*. EDP Sciences, pp. 311–336.

Patents

Poli, Jean-Philippe, Laurence Boudet, Bruno Espinosa, and Laurence Cornez (2019). “Accesneur Sémantique Aux Données”.

Poli, Jean-Philippe (Oct. 2016). “Software Architecture for Expert System”. US Patent App. 15/516,859.

Aupetit, Michael, Ricardo De Aldama, Jean-Philippe Poli, and Laurence Boudet (2015). “Dispositif Et Procédé D’optimisation D’un Système D’inference Floue Préservant L’interprétabilité”.

Poli, Jean-Philippe and Jean-Paul Laurent (Nov. 2015a). “Assisted Input of Rules into a Knowledge Base”. US Patent App. 14/618,350.

Auribault, Karine, Jean-Denis Muller, Géraldine Cancel-Tassin, Olivier Cussenot, Stéphane Gazut, Nicolas Gilardi, David Mercier, Jean-Philippe Poli, Emmanuel Ramasso, and Frédéric Suard (Dec. 2011). “Prediction Method for the Screening, Prognosis, Diagnosis or Therapeutic Response of Prostate Cancer, and Device for Implementing Said Method”. US Patent App. 13/056,746.

PhD

Poli, Jean-Philippe (2007). “Structuration Automatique de Flux télévisuels”. PhD thesis. Université Paul Cézanne.

Softwares

Jean-Philippe Poli et al. (from 2008). *ExpressIF*. under license.
— (from 2010). *PACT / ExpressIF sensors*. under license.

Appendix G

Selection of articles

In this annex, I gathered 9 articles that, according to me, represent my work.

- Hotel, Olivier, Jean-Philippe Poli, Christine Mer-Calfati, Emmanuel Scorsone, and Samuel Saada (2017a). “Estimation of the Parameters of Saw Sensor’s Frequency Shift: Application to Odour Recognition and Concentration Evaluation”. In: *2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN)*. IEEE, pp. 1–3.
- Poli, Jean-Philippe and Laurence Boudet (2018). “A Fuzzy Expert System Architecture For Data And Event Stream Processing”. In: *Fuzzy Sets and Systems* 343, pp. 20–34.
- Grivet Sébert, Arnaud and Jean-Philippe Poli (2018a). “Material Classification from Imprecise Chemical Composition: Probabilistic vs Possibilistic Approach”. In: *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Cherrier, Noëlie, Jean-Philippe Poli, Maxime Defurne, and Franck Sabatié (2019). “Consistent Feature Construction with Constrained Genetic Programming for Experimental Physics”. In: *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, pp. 1650–1658.
- Baaï, Ismaïl and Jean-Philippe Poli (2019). “Natural Language Generation of Explanations of Fuzzy Inference Decisions”. In: *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Poli, Jean-Philippe, Wassila Ouerdane, and Régis Pierrard (2021a). “Generation of Textual Explanations in XAI: the Case of Semantic Annotation”. In: *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Pierrard, Régis, Jean-Philippe Poli, and Céline Hudelot (2021). “Spatial relation learning for explainable image classification and annotation in critical applications”. In: *Artificial Intelligence* 292, p. 103434. ISSN: 0004-3702.
- Hajri, Hiba, Jean-Philippe Poli, and Laurence Boudet (2021b). “Towards Monotonous Functions Approximation from Few Data With Gradual Generalized Modus Ponens: Application to Materials Science”. In: *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 796–800.
- Kunitomo-Jacquín, Lucie, Aurore Lomet, and Jean-Philippe Poli (2022). “Causal discovery for fuzzy rule learning”. In: *2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.