Hierarchical Decision Granules Optimization through the Principle of Justifiable Granularity

Raúl Navarro-Almanza, Mauricio A. Sanchez, Juan R. Castro,
Olivia Mendoza, Guillermo Licea

Universidad Autónoma de Baja California,
Facultad de Ciencias Químicas e Ingeniería,
Mexico

{rnavarro, mauricio.sanchez, jrcastror, omendoza, glicea}@uabc.edu.mx

Abstract. Interpretable Machine Learning (IML) aims to establish more transparent decision processes where the human can understand the reason behind the models' decisions. In this work a methodology to create intrinsically interpretable models based on fuzzy rules is proposed. There is a selection to identify the rule structure by extracting the most significant elements from a decision tree by the principle of justifiable granularity. There are defined hierarchical decision granules and their quality metrics. The proposal is evaluated with ten publicly available datasets for classification tasks. It is shown that through the principle of justified granularity, rule-based models can be greatly compressed through their fuzzy representation, not only without significantly losing performance but even with compression of 40% it manages to exceed the performance of the initial model.

Keywords. Granular computing, neuro-fuzzy, Sugeno, hierarchical decision granules, interpretable machine learning.

1 Introduction

Machine learning models are ubiquitous nowadays; They are involved in many activities of daily life in which people are aware or unaware of their use. It is essential to know how and why the models provide a particular output and how they can be made more fair and secure, especially in critical applications. Interpretable Machine Learning (IML) strives to construct a bridge between the learned model and human understanding.

There are various ways to achieve model interpretability by applying: i) intrinsic interpretable models, ii) regularization techniques, and iii) posthoc explanation techniques. One trend is to build surrogate models (lower complexity) than the original one and more interpretable to understand the decision process, such as rule-based models. [7, 9, 48].

This work takes advantage of the inherent interpretability that brings Fuzzy Inference Systems (FIS). FIS models are interpretable since they are rule-based models and are described linguistically. The antecedent is represented by fuzzy variables and sets, proposed by Zadeh [59]. The inference process of those systems is performed through fuzzy reasoning, which aims to represent human perception and their inference mechanism under uncertainty.

An interesting characteristic of FIS is that their knowledge representation is composed of IF-THEN rules, where their antecedents and consequents are in natural language. For example, a proposition can be "`Temperature` is `hot`", where `Temperature` is related to an input attribute and `hot` to the set which partially belong. As was described, this formal notation potentially brings an intrinsic high interpretability degree if their components are well-defined [32].

The FIS are used in a variety of application domains in ML context, such as medical [14, 19, 24, 41, 57, 35], robotics [15, 1], decision making [60, 12, 3].

Often this FIS modeling is data-driven, usually by some unsupervised technique (e.g., clustering) to discover their input partition, membership functions parameters, and rule structure. An important issue is that as the number of attributes (input dimension) and fuzzy sets increases, it becomes more susceptible to presenting a combinatorial problem in the rule-finding process. In this approach, we construct the initial rule structure through a decision tree model, which is further transformed to the fuzzy space domain and adapted to a Sugeno-type architecture described in the further sections.

As the input dimension space and feature interaction complexity increases, the resulted decision tree gets deeper; therefore, it becomes pruned to be hardly interpretable. For each *leaf* node or decision node in the tree, it can extract an IF-THEN decision rule. To tackle the problem of a high number of decision rules is conducted a post-pruning process. The pruning process is carried on by selecting the most relevant elements (antecedents and decision rules).

Granular Computing (GrC) aims to form meaningful Information Granules, which represent a collection of objects abstraction and relate them by some similarity in a hierarchical manner [5], which allows creating semantically richer structures [67]. GrC is inspired by how the human brain works, processing information abstractly at the required level to resolve a given task. The data is organized by some of their characteristics in a hierarchical way to form a granule, a formal representation of this structure with two essential properties, specificity, and coverage.

Specificity is related to the granule representation; the higher specificity is, the less ambiguous it is, and humans can easily understand it. On the other hand, coverage is related to the proportion of individuals designated by the granule.

Intuitively, a well-formed information granule should have higher values in both specificity and coverage. However, these properties are commonly in conflict, as the higher the specificity is, the lower is the coverage; an optimization process usually conducts the granule allocation.

This work proposes a data-driven method to construct a fuzzy rule-based system using the principle of justifiable granularity for selecting the most relevant knowledge base elements, according to the trade-off between specificity and coverage values. This discrimination is conducted over the rules formed by decision tree models, allowing the construction of a variable model complexity useful in IML.

The main contributions of this work are:

— Definition of a Sugeno-type neuro-fuzzy model for classification tasks that leads to the straightforward interpretation of the decision process.

— Characterization of hierarchical Information Granules in decision-set context.

— Definition of hierarchical specificity and coverage metrics for optimization of graph-based entities.

— Data-driven methodology for establishing fuzzy inference system structure by decision tree rule extraction and selecting the most relevant elements by following the principle of justifiable granularity.

In the following sections correspond to: a brief description of the relevant theory, section 2. In section 3 is reviewed the related work. In section 4 is described the hierarchical Information Granules and their generation. In section 5, the neuro-fuzzy model is described. In the section 6 is shown the followed methodology and experiment setup to evaluate the proposal. The results and conclutions are in sections 7 and 8 respectively.

2 Background

The proposed methodology comprehends three main areas: i) Fuzzy systems for establishing natural language interface in the form of fuzzy rules; ii) Decision tree-based models for initial structure rule discovery in the data space domain. iii) Granular computing, aiming at the well-formed Information Granule for rule-based knowledge model representation following the principle of justifiable granularity.

2.1 Fuzzy Systems

Fuzzy Systems are rule-based models that use fuzzy logic to conduct the reasoning process. Fuzzy logic was proposed by Zadeh [58] as an approach to represent computable human perception through words.

This reasoning system type offers greater explainability and is widely used in the Machine Learning field due to its ability to process linguistic information [13].

These systems are broadly used on various domain applications such as control, medical, aerospace and environmental applications [15, 30, 54, 49, 29, 9].

The knowledge base modeling can be built either manually by experts, or designed automatically, usually by clustering techniques.

Zadeh's fuzzy rule has the following structure: IF *Temperature* is *hot* THEN *Cooling* is *high*.

The antecedents and consequents are as shown formed by linguistic variables (*Temperature* and *Cooling*), the values of these variables are linguistic values (*hot* and *high*), which their meanings are easily understanding by humans.

The Zadeh's linguistic variable [59] is characterized by a quintuple $(x, T(x), X, G, M)$, in which x is the name of the variable; $T(x)$ is the term set of x, linguistic terms; X is the universe of discourse; G is a synthetic rule which generates linguistic terms in $T(x)$; M is a semantic rule which associates each linguistic value A its meaning $M(A)$, where $M(A)$ denotes a fuzzy set in A.

A fuzzy set A in X domain is defined as a set of ordered pairs (equation 1):

$$A = \{(x, \mu_A(x)) | x \in X\}, \tag{1}$$

where $\mu_A(x) \in [0, 1]$ is the membership function that represents the human perception in form of membership degree in de universe of discourse X.

2.2 Decision Trees

Decision trees are graph-based Machine Learning models for classification and regression tasks. The model constructs a tree in which their non-terminal nodes perform splits in the input data space (in the context of Machine Learning). The splitting process is sequential until it reaches the leaf nodes (terminal nodes). The evidence provides an output label or probability of belongingness to some class (in classification tasks).

This tree construction relies on subsequent partitioning in the data inputs space by selecting the best feature and value to split. There are many criteria for select the best split candidate. The main idea is to achieve the best purity, which means that the residual data after each split belongs to only one class; until this goal is not reached, new nodes are added to the tree.

The relevant hyper-parameters in this context for regularizing the model are the number of features searched in each partition node; minimum elements belonging to a node to consider creating a branch; criteria to measure the quality of a split; the maximum depth that can have the decision tree.

Once the tree is created, it can be traverse through its branches until each lead node is reached; every node condition in the path can be extracted to form an antecedent and consequent part of an IF-THEN rule. Thus, for every leaf node, a rule can be constructed.

In this work, the performed input space partitioning by the decision tree model is used to define the initial structure of the knowledge base of the FIS. The selection of the most relevant rules elements is conducted via optimization by the principle of justifiable granularity.

2.3 Granular Computing

Granular Computing is a paradigm inspired by how the human brain performs different levels of entities' characteristics abstraction and uses those to make decisions. The main particle in this paradigm is called *Information Granule*. These granules can be regarded as a collection of objects hierarchically that exhibit similarities among them.

There is not an specific formalization to define an Information Granule, they can be described by a huge variety of different representation, such as: interval sets [21, 40], rough sets [26, 61, 50, 49, 69], fuzzy sets [37, 4, 36, 63, 62], probabilistic sets [42, 53], possibility sets [65, 46], neural networks [34, 20, 64, 51, 28, 17]. GrC is a unified framework of techniques, methodologies, and theories for the formalization, construction, and manipulation of Information Granules; it brings a coherent environment to work with abstract object representation [5].

Some techniques for building fuzzy information representations are inspired by granular computing, such in [43] where a method is proposed to find the right cluster size concerning the data context; in [8] is proposed a generalized Type-2 fuzzy control model that uses granularity to divide the global model by simpler models.

2.3.1 Principle of Justifiable Granularity

The fundamental idea of principle of justifiable granularity is to form meaninful Information Granules based on experimental evidence (data), following two general criterias: *coverage* and *specificity* [38].

The coverage is the numeric evidence that supports the Information Granule. The intention is to form/discover granules with the more substantial experimental evidence that supports its formulation. On the contrary, the specificity is related to the granule's well-formed; the smaller the Information Granule is, the better. The ideal is to form meaningful Information Granules with the higher coverage and specificity as possible.

These two requirements are in conflict. In a basic formulation, the granule A been represented as an interval $[a, b]$. As higher the range is, the more expected experimental support (cardinality, showed as $card(\cdot)$) it gets ($cov(A) = card(\{x_i | x_i \in [a, b]\})$); at the same time, the specificity decreases, considering the range as the specificity ($sp(A) = |b - a|$).

To find the best meaningful Information Granule, this contrary behavior between the criteria of coverage and specificity can be defined as a multi-objective optimization problem for the maximization of the composite multiplicative index. Given a set of design parameters θ for the Information Granules, it must find the best values for θ that maximizes the equation 2:

$$A^*_\theta = \underset{\theta}{\operatorname{argmax}} \quad \text{cov}(A_\theta) * \text{sp}(A_\theta). \tag{2}$$

The principle of justifiable granularity allows finding the best well-formed granule. In the fuzzy logic context, it has been used to define fuzzy information granules, such in [33] where is used to construct IT2 Fuzzy Memberships functions.

The proposed method in this work applies the principle of justifiable granularity to compress decision sets and improve their interpretability.

3 Related Work

GrC has been used in Machine Learning problems as a way to define semantic richer data representation to build models with missing information [25], prototype forming for descriptors of facial expressions [55]. To establish initial neural network architecture for further optimization [39]. Furthermore, this framework has been used to overcome the limitations of existing Machine Learning models related to data quality [22, 10, 6], interpretability, domain adaptation for regression tasks [22], and dimensionality reduction [2, 16, 52].

In [31] is discussed the importance of adopting the GrC paradigm in rule-based systems as a way to improve interpretability. The principle of justifiable granularity has been used to discover robust information clusters in the context of data-driven system modeling [68].

There are various works in the context of GrC which support the hypothesis of more robust generation rule-based systems [44, 54], rule reduction using complex fuzzy measures with GrC [47]. Also, the hierarchical representation of granule modeling has been used to solve hierarchical classification problems [23], for building interpretable models in data stream learning environments [27].

The use of GrC in the context of Machine Learning comprehends the discovery of the Information Granules in the data space domain and forms them by some formal description, e.g., intervals, fuzzy sets, rough sets, hyperboxes.

Some recent approaches in GrC adopt cognitive science perspectives and fuzzy logic to support intelligent decision-making [18, 56]. GrC has promoted the adoption of fuzzy logic for data abstraction to be capable of processing various data types in classification tasks using granular decision trees [29, 30]. A top-rated operator in deep learning for extracting relevant features, the convolution, had been adapted to operate with fuzzy sets with a granular perspective for classification problems [11]. In [66] is proposed a polynomial-feature granulation method based on long short-term memory network for oxygen supply network prediction.

In this work, a decision-tree model discovers the rule base before characterizing its elements in a granular paradigm. Given an a priori defined granules collection, this work selects the best ones to create higher-level Information Granules. It then performs the fuzzification to develop a new Sugeno-type fuzzy rule base with learning capability due to their analogous neural network representation.

4 Granules Construction

Different levels of abstractions define the granule construction process. The level of these granules is denoted by the subindex A_i. In the first level of granularity, A_0 corresponds to the original data space, so that a zero-level granule is equivalent to a given instance of the dataset $A_0 \approx x^{(i)}$.

A level 1 Information Granule (A_1) is characterized as the tuple (m, r) that correspond to a range $[m - r, m + r]$, notice that in this particular scenario, it is defined with a symmetric proportion from the median (could be further extended). A level 2 Information Granule (A_2) is described as an implication relation; at this abstraction level, the interaction between different domain Information Granules is considered. A level 3 Information Granule in this work is defined as a decision set.

The notation to denote a granulation process is through eq. 3:

$$\mathcal{G}(\mathcal{A}'_i) = A_{i+1}. \tag{3}$$

where $\mathcal{G}$ is the mapping process to form a higher Information Granule given a set of lower-level granules, $\mathcal{A}'_i \subseteq \mathcal{A}_i$ denote a set that belongs to the i-level granular space, and A_{i+1} is a formed higher level granule. For instance, the first abstraction level starts from the crisp data space, such that $\mathcal{G}(X) = \mathcal{G}(\mathcal{A}_0) = A_1$. For notation purposes $\mathcal{G}^l(A)$ denotes the process of perform l-level abstraction processes for a given granule A.

The notation to denote a degranulation process is through eq. 4:

$$\mathcal{G}^{-1}(A_i) = \{A_{i-1}^{(1)}, \ldots, A_{i-1}^{(n)}\} \subseteq \mathcal{A}_{i-1}, \tag{4}$$

where $\mathcal{G}^{-1}$ is the mapping process to form a lower Information Granule given a granule, $\mathcal{A}_i$ denote the i-level granular space, and A_{i-1} is a formed lower level granule.

It is essential to notice that the raw representation of a granule is a set that is formed by granules of lower levels. Intuitively, a granule should be represented by a model that requires low information.

For a level 1 Information Granule is defined the following metrics to measure the coverage (eq. 5) and specificity (eq. 6).

Coverage:

$$cov(A_1) = \frac{card(\{x_k | (m - r) < x_k < (m + r)\})}{N}. \tag{5}$$

Specificity:

$$sp(A_1) = 1 - \frac{|r|}{|X_{max} - m|}, \tag{6}$$

where A_1 is an Information Granule, $x_k \in X$, m, and r characterize a range (level 1 granule); m is the median of the range, and r the distance to the rage limits in a symmetric way. $card(\cdot)$ stands for the cardinality of a given set:

$$\mathcal{G}(A'_1) = \bigwedge_i \rho({A_1}^{(i)}; \mathbf{x}) \rightarrow Y = A_2, \tag{7}$$

where $A'_1 \subseteq \mathcal{A}_1$, that in their raw representation is a set that belongs to the $\mathcal{A}_1$ space, namely a set of ranges (denoted by a tuple (m, r)) which conforms an IF-THEN rule.

The operator $\wedge$ represent the conjuntion operation. $\rho : \mathcal{A}_1 \times X$ is a logical function that forms the proposition "x `belongs to the granule` $\mathcal{A}_1^{(i)}$". Y is the target domain:

$$\mathcal{G}^{-1}(A_2) = \mathcal{A}'_1, \quad \mathcal{A}'_1 \subseteq \mathcal{A}_1. \tag{8}$$

In the degranulation process of a level 2 granule, the raw representation of the operation is a set of level 1 granules, and those granules form the antecedent part in the rule structure.

For a level 2 Information Granule are defined the following metrics to measure the coverage (eq. 9) and specificity (eq. 10).

Coverage:

$$\begin{aligned} cov(A_2) = & \frac{1}{N} \sum_{i}^{N} card(\{x^{(i)} | \forall x^{(i)} \in X, \\ & \forall A_1^{(i)} \in \mathcal{G}^{-1}(A_2). \rho(A_1^{(i)}; x^{(i)})\}). \end{aligned} \tag{9}$$

Specificity:

$$sp(A_2) = \frac{card(\{\mathcal{G}^{-1}(A_2)\})}{dim(X^{(i)})}, \tag{10}$$

where A_2 is an Information Granule formed by a implication relationship $\bigwedge_i \rho({A_1}^{(i)}; \mathbf{x}) \rightarrow Y$, N is the number of instances that correspond to the dataset; and, $dim(X^{(i)})$ is the number of considered features in the dataset. $card(\cdot)$ stands for the cardinality of a given set.

To select a justifiable Information Granule, it is necessary a measure their formation quality. Due to the contradictory behavior of coverage and specificity can form a Pareto front to select the best candidates. Notice that the Pareto front is formed by the product and can be computed to any granule at any abstraction level (eq. 11):

$$Q_{l_i}(A_i) = sp(A_i) * cov(A_i)^{\gamma_i}. \tag{11}$$

There is a parameter that serves to prioritize one of the terms [38], γ_i. If $0 \geq gamma_i < 1$ there is pondered more the coverage, on the contrary, if $\gamma_i > 1$ then the specificity gets more relevant in the calculus. For each abstraction level, a different value γ_i can be defined.

4.1 Hierarchical Quality Measurement

The quality construction of a given of H level granule is measure with the proposed metric (eq. 12), which perform successive degranulation process and multiply their Pareto front values to the lower level granules:

$$V(A_h) = \{Q_{l_h}(A_h) \times V(A'_{h-1}) | \forall A'_{h-1} \in \mathcal{G}^{-1}(A_h)\}. \tag{12}$$

The optimization problem is shown in equation 13, which aims to find the more appropriate level 1 granules. The aptitude function is the hierarchical measure (eq. 12), their restrictions are: 1) The solution set must be the minimum cardinality (showed as the function $card(\cdot)$) 2) The aggregation of the values v_i (hierarchical Pareto) should be equal or less the to regularization parameter α, and 3) The number of elements in the solution set must be equal or less than the regularization parameter ξ. These two regularization parameters allow finding the best level 1 granules smaller set with at least a cumulative value of α and does not have more than ξ elements:

$$\begin{aligned} A_H^* = & \underset{\mathcal{A}'_1 \subseteq \cdots \subseteq \mathcal{A}_H}{\operatorname{argmax}} \; V(\mathcal{G}^2(\mathcal{A}'_1)) \\ & \text{subject to:} \\ & 1) \quad \min card(\{\mathcal{A}'_1\}), \\ & 2) \quad \sum_{i=1}^{card(\{\mathcal{G}^{-1}(\mathcal{A}_H)\})} V(\mathcal{G}(\mathcal{A}'_1))_i \leq \alpha, \\ & 3) \quad card(\{\mathcal{A}'_1\}) \leq \xi. \end{aligned} \tag{13}$$

The resulted Information Granule collection reconstructs the decision set, formally by $G^2(A^*)$. This process can be treated as a post-pruning technique, the bias of the model increases while its variance decreases. The initial structure for building the knowledge base of the FIS to be optimized is formed by the decision set. Figure 1 shows the general block diagram of the proposed methodology for data-driven fuzzy rule base construction. The details of the neuro-fuzzy model characteristics are in section 5.

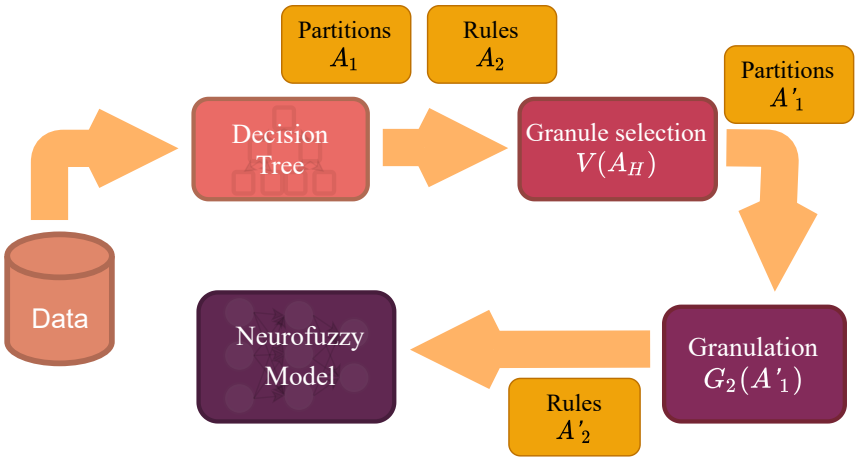

Fig. 1. Block diagram of the proposed methodology for building a fuzzy rule-based system from decision set created by decision-tree model using the principle of justifiable granularity

5 Sugeno-Type Neuro-Fuzzy Model

The Sugeno fuzzy systems allow their construction in a systematic form to generating fuzzy rules in a data-driven manner, they were proposed by Takagi, Sugeno, and Kang [45]. These systems are also composed of IF-THEN rules, but only their antecedent belongs to the fuzzy space. The consequent part is a crisp function that maps the input space.

As in the Mandani type fuzzy systems, the rules might fire all at once to get a crisp output value, can compute a simple weighted average of function outputs, which is less time-consuming than defuzzification in Mamdani type fuzzy systems. The knowledge base can be described as follows:

$$R^1: \text{IF } x_1 \text{ is } low \text{ and } \ldots \text{ and } x_m \text{ is } low \text{ THEN, } y \text{ is } \sigma^{(1)}_{j\in J}(\mathbf{x};\mathbf{w}),$$
$$R^2: \text{IF } x_1 \text{ is } low \text{ and } \ldots \text{ and } x_m \text{ is } high \text{ THEN, } y \text{ is } \sigma^{(2)}_{j\in J}(\mathbf{x};\mathbf{w}),$$
$$\vdots$$
$$R^n: \text{IF } x_1 \text{ is } high \text{ and } \ldots \text{ and } x_m \text{ is } low \text{ THEN, } y \text{ is } \sigma^{(n)}_{j\in J}(\mathbf{x};\mathbf{w}),$$

where x_i is a fuzzy variable (which models the feature space), their fuzzy values are represented by the terms *low*, *high*, etc. y is the output space described by the function $\sigma^{(n)}_{j\in J}(\mathbf{x};\mathbf{w})$, where J represents the output classes, $\mathbf{x}$ are the input crisp values, and $\mathbf{w}$ are the function's coefficients.

Once the knowledge base is designed, some optimization methods can adjust their membership function parameters to fit the data better. This optimization process comprehends the membership function and consequent crisp function parameters. In this approach, to maintain the interpretability characteristic, sigmoid and linear functions are selected.

An analogous neuro-fuzzy architecture carries out the optimization of fuzzy model parameters. This architecture comprises five layers: input, fuzzification, inference, implication, and de-fuzzification layer. The connections between the layers are not fully connected to maintain coherent antecedent relationships in the fuzzy rules. In the fuzzification layer are only connected the membership functions belonging to the input domain. In the implication layer, each crisp function is only related to the rule's output (a given class). Figure 2 shows a visual representation of a Sugeno-type neuro-fuzzy model.

The parameters to optimize the model belong to the antecedent part of the fuzzy rule and the (trainable) parameters in the crip consequent function.

The neural architecture is shown in figure 2. The first layer is non-fully connected among neurons that represent the fuzzification process. Wich maps a crisp input to the fuzzy space:

$$f^{B^j_k}(x_i) = \mu_{B^j_k}(x_i), \tag{14}$$

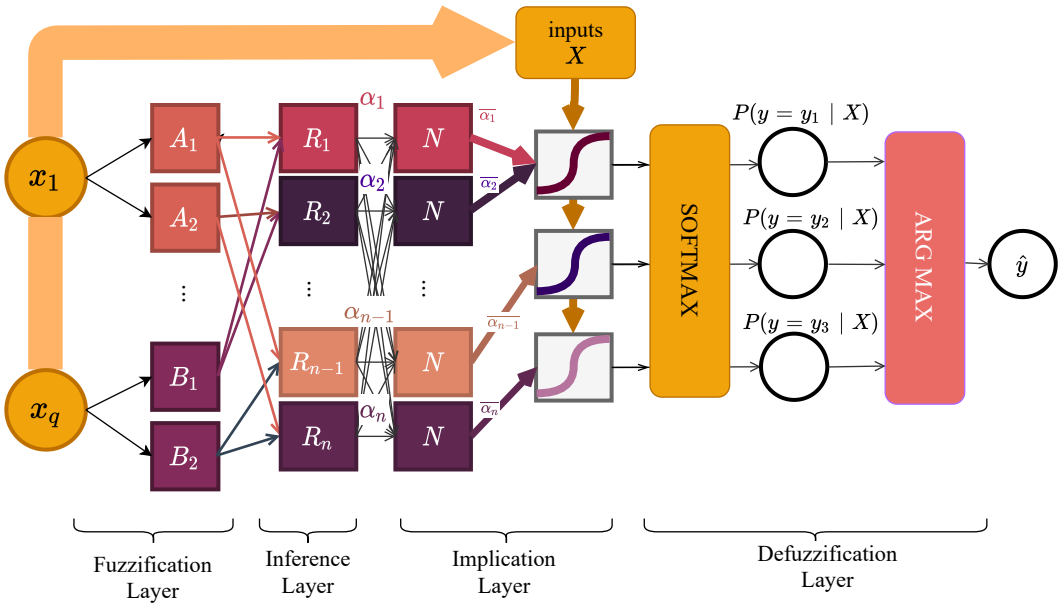

Fig. 2. The Sugeno Neuro-fuzzy representation is composed of a non-fully connected 5-layer artificial neural network. All the computations correspond to the involved operations in a Sugeno-type fuzzy inference system for q inputs, n rules, and crisp consequent functions. Each rule is only linked to one class $j \in J$ target space

where $B_r^j \in V_k$ is a fuzzy set that belongs to the fuzzy variable V_k, the domain of each fuzzy variable is shared by its corresponding attribute domain in the dataset. Only the membership functions directly related to the attribute are evaluated by the input value, which results in a semi-connected layer.

The inference layer is also non-fully connected, fires at a certain strength value in the range $[0, 1]$. The implication operation calculates a *t-norm* ($\tilde{*}$) as a product:

$$\alpha^l(x_i) = \prod_{r=1}^{p} f^{B_r}(x_i). \tag{15}$$

The implication layer performs a normalization operation that conforms a step to the weighted average to the output:

$$\bar{\alpha}^l(x_i) = \frac{\alpha^l(x_i)}{\sum_{j=1}^{L} \alpha^j(x_i)}. \tag{16}$$

After the normalization process, for each rule that has a crisp function consequent related to an output class $j \in J$, a $\tilde{*}$ as the product is calculated:

$$z^j(x_i) = \sum_{l=1}^{M} \{\sigma(x_i; \mathbf{w}^l) \times \bar{\alpha}^l(x_i)\}, \tag{17}$$

where σ corresponds to sigmoid function that transform the input space, $\sigma(x_i; \mathbf{w}^l) = \frac{1}{1+e^{-\mathbf{w}^l x_i}}$. After the sigmoid transformation, all values are computed by the softmax function, which maps proportionally for each class, values in the range $[0, 1]$ and $\sum_{j=1}^{J} h_j = 1$, it commonly interprets those values as a probability output:

$$h_j = P(y = j|\mathbf{x}) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}. \tag{18}$$

The output value dimension corresponds to the number of classes in the problem domain. Getting an actual label as a result value could be computed just as the argument position with the higher probability value:

$$\hat{y} = \arg\max_{j \in J} P(y = j|\mathbf{x}). \tag{19}$$

This neuro-fuzzy model adjusts its parameters to fit a given target better. To measure the error of the prediction is used the Cross-Entropy loss function ($Loss_{CE}$):

$$Loss_{CE}(\mathbf{y}, h(\mathbf{x}; \mathbf{w})) = \\ -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J} y_j^{(i)} \log(h_j(x^{(i)})) + \lambda_1||\mathbf{w}||_1 + \lambda_2||\mathbf{w}||_2, \tag{20}$$

where N is the number of instances on the batch; y is the target value in the form of `one-hot-encoding`, and h_j is the predicted probability of belonging to the class j. In the loss function are defined regularization l_1 and l_2 norms to constrain the weight values and prevent overfitting. These regularizers can help to improve the interpretability by only *relevant* considering relevant features.

The trainable parameters on the neuro-fuzzy model are the design parameters of the fuzzy sets and the set of crisp consequent function parameters; in this setup, the membership functions are defined by Gaussian functions, then the trainable parameters are the mean and σ values, where $\sigma > 0$.

A gradient descent optimization method is applied to find the best parameters. The learning rule is shown in equation 21:

$$\theta^{\text{new}} = \theta^{\text{old}} - \eta \nabla E(\theta^{\text{old}}), \tag{21}$$

where θ are the parameter vector values; $E(\theta)$ is the gradient of the error value of the model with the parameters θ; η is the learning rate value in $0 < \eta < 1$.

The hyper-parameters of the model are:

— *Learning rate value:* this value scales the directional vector generated by gradient calculation, as the lower the value, the better search is but slower. Usually, the default value is set to a value of 0.01.

— *Batch size:* the selection of dataset partition to train the model is set by this value.

— *Epoch number:* an epoch represents an entire iteration overall dataset (train dataset partition).

— *Goal error value:* is a threshold value to consider to stop the training process because is considered acceptable at that value.

— *Output function structure:* is the computation that transforms the input value to a crisp output value to further averaging pondered by the firing strength values.

— *Regularization coefficients λ_1 and λ_2:* those restrict how much the trainable parameters in the output crips functions increases.

6 Experimentation

Ten publicly available datasets [1] are used in order to evaluate the proposed model under different domain applications. All datasets correspond to classification tasks; their characteristics are shown in the table 1.

Table 1. The selected publicly available datasets at UC Irvine Machine Learning Repository[1] for evaluating the proposed methodology

	dataset	instances	features	classes
1	abalone	4177	8	3
2	credit-g	1000	20	2
3	creditcard	284807	29	2
4	diabetes	768	8	2
5	ionosphere	351	34	2
6	iris	150	4	3
7	sonar	208	60	2
8	spambase	4601	57	2
9	wdbc	569	30	2
10	wine	178	13	3

The methodology consists of primary three steps:

1. Decision tree construction to extract and generate a rule-based decision set.
2. The decision set reconstruction following the principle of justifiable granularity for the selection of the more meaningful granules.

[1] https://archive.ics.uci.edu/ml/index.php

3. The construction of Sugeno-type neuro-fuzzy architecture for classification tasks then optimized their membership function design parameters and coefficients of the consequent output functions.

6.1 Decision Tree Construction

Given a training dataset $\mathcal{D}_{train} = \{(x, y)\}_{i=1}^{N}$, a decision tree model is train to map the input data patterns to the target domain space, $f_{tree}(\mathbf{x}; \mathrm{T}) \rightarrow \mathbf{y}$. The hyper-parameters set for this experiment are: i) complete search of the feature space in each partition split; ii) one element at minimum belonging to a node to create a branch; iii) Gini impurity criteria to measure the quality of a split; iv) without the maximum depth of the three, that means the nodes are expanded until all leaves are pure.

Creation of a decision set (M_{ds}) by traversing the tree paths from the root to the leaves nodes (decision nodes). Due to the possible repetition of some features for the splitting, it is necessary to simplify the rules by limiting each feature to be clustered only in one range (this process maintains the model's fidelity and does not affect the original representation outcome). Each leaf node creates a rule; therefore, it can be a potentially large number of them. Some criteria must clip all feature ranges. In this experiment, the maximum and minimum values for each feature are taken to replace those undefined boundaries. At this step, every node has been contributed to creating intervals $[m, r]$ where m is the mean value and r is the distance from the mean to the left and right, that are further used to form proposition such as "x `is in` $[m - r, m + r]$".

6.2 The Decision Set Reconstruction by the Principle of Justifiable Granularity

The intervals formed from the learned tree T compounds the first level Information Granules. All ranges created by the decision tree are represented with the tuple (m, r), where m is the median of the range, and r is the distance to some boundary (notice that only represents symmetrical granules).

Once all level 1 Information Granules are generated ($\mathcal{A}_1$ space), then by following the antecedents of the rules, the level 2 Information Granules are generated ($\mathcal{A}_2$ space). At this level, the relationship between lower-level granules are established (see section 4). Next, those level 2 Information Granules are grouped to form a structure-less level 3 Information Granule; namely, it represents the decision set (M_{ds}).

Due to the potentially large number of elements in the decision set M_{ds} (now forming a level 3 granule), it is necessary to prune it. The pruning process follows the principle of justifiable granularity; in this context, instead of selecting a numerical range of values, the best set of level 1 Information Granules that carry out more meaningful information in terms of coverage and specificity. Next, the optimization process (defined in the section 4) is conducted to find the best lower level granules ($\mathcal{A}_1^*$) that contribute the most to create higher quality Information Granules of higher levels.

The set of level 1 granules are clustered by the granulation process to reconstruct a decisition set with fewer information (less variance and more bias) since the lack of some elements (antecedents and rules), $M'_{ds} = \mathcal{G}^2(A_1^*)$, where $M'_{ds} \subseteq M_{ds}$.

6.3 Sugeno-Type Neuro-Fuzzy Optimization

Once the decision set is reconstructed (M'_{ds}), it is converted to a fuzzy inference system using Gaussian membership functions to represent the antecedent ranges($[m - r, m + r]$). In other words, the crisp Information Granules A_1 are fuzzified. Then those are defined by a tuple (m, σ) to represent the Gaussian membership function, where m is the same as the mean of the range in the initial A_1, and σ is the standard deviation, the value approximates this value r, such that $\sigma = r/2$.

The consequent part of each rule is used to define a sigmoid function that is related to the target class $\sigma_{j \in J}^{(r)}(\mathbf{x}; \mathbf{w})$, where r denotes the rule and j the specific class. This sigmoid computes the dot product between the input vector and a set of initially random weights $\sigma(\mathbf{w}^T\mathbf{x})$; each class at least has a function associated with it. The output of this function evaluation is multiplied by

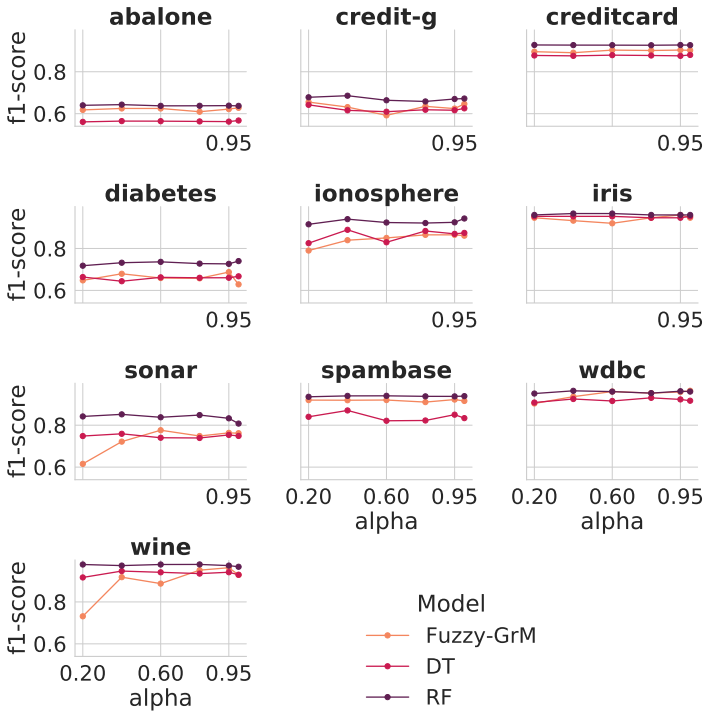

Fig. 3. Performance comparison between the decision tree-based models (Decision Tree and Random Forest) and the proposed granular model (Fuzzy-GrM) at diverse α values. As lower the α value, the higher compression the model is

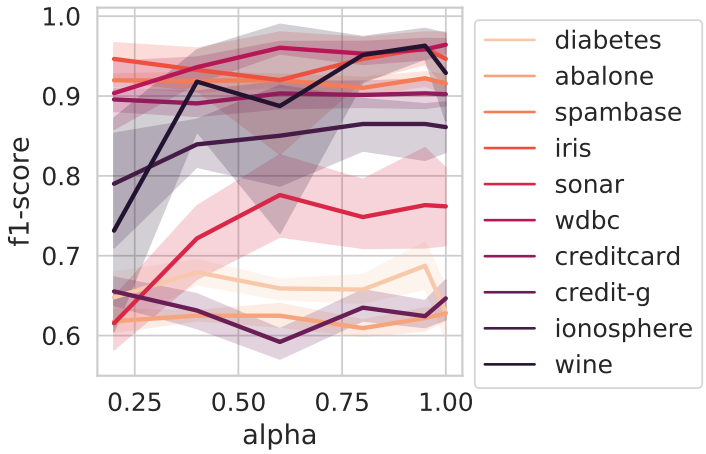

Fig. 4. Overall performance comparison with diverse values of alpha, applied in the ten different classification datasets

Table 2. The proposed model average results with different alpha values applied in 10 publicly available datasets

	alpha	**0.20**	**0.40**	**0.60**	**0.80**	**0.95**	**1.00**
DT	mean	0.79	0.80	0.79	0.80	0.80	0.80
	σ	0.13	0.15	0.14	0.14	0.14	0.14
RF	mean	0.85	0.86	0.86	0.85	0.86	0.86
	σ	0.13	0.13	0.13	0.13	0.13	0.13
Fuzzy-GrM	mean	0.77	0.81	0.81	0.82	0.83	0.82
	σ	0.14	0.13	0.15	0.14	0.14	0.14
Element reduction	mean	93.91%	87.21%	81.02%	71.20%	58.61%	49.09%
Rule reduction	mean	82.51%	71.61%	61.38%	49.09%	41.10%	39.48%

the normalized firing strength of the rules that are linked to the target class consequent (equation 16). At this point, some selection criterion defines the output (e.g., the label of a more significant output rule).

In this work is used a softmax computation, to define the probabilistic output of the FIS (equation 18). In addition, to better interpret inputs, it creates a smooth solution surface space in the training step of the neuro-fuzzy by gradient descent-based optimization algorithms.

7 Results

The proposed model was evaluated by 5-fold cross-validation, in 10 publicly available datasets for classification, with 6 different values for the hyper-parameter α which serve up to select the compression level by selecting the most relevant elements (according to eq. 12). The obtained results for overall Sugeno-type neuro-fuzzy model for classification (showed in table 2) were f1-scores of 0.77, 0.81, 0.81, 0.82, 0.83, 0.82 when α-values were set to 0.2, 0.4, 0.6, 0.8, 0.95, 1 respectively, with the maximimun rules parameter (ξ) set to 50.

Considering all different values of α, in average the model reduction (in terms of elements) was 73.51% with $\sigma = 17.23$. From a rule percentage reduction perspective, the average compression was 57.53% with $\sigma = 17.34$, and relative error concerning the Simple Decision Tree model of -1.4% with $\sigma = 0.021$, and 5.55% with $\sigma = 0.022$ respect to Random Forest model (in this experiment the number of weak learners was set to 100).

Figure 3 shows the performance comparison between the decision tree-based models and the proposed one at diverse α values.

At the global level, considering all analyzed datasets, the minimum elements compression percentage given the following values of the prunning value α were: 81.25% with $\alpha = 0.2$, 58.33% with $\alpha = 0.4$, 50% with $\alpha = 0.6$, 23.07% with $\alpha = 0.8$, 81.25% with $\alpha = 0.2$, 11.11% with $\alpha = 0.95$, 0% with $\alpha = 1$. At rule compression percentage were obtained: 57.14% with $\alpha = 0.2$, 33.33% with $\alpha = 0.4$, 25% with $\alpha = 0.6$, and no rule compression at all for higher values for α.

In dataset results tables (10-7), there are comparisons of the f1-score of the proposed model (F-GrM), Decision Tree (DT), and Random Forest (RF). Figure 5 shows as higher compression is applied to the model, the higher the variance is. Figure 4 shows the overall performance comparison of different hyperparameters.

A paired sample t-test over the mean f1-score values was used to formally validate the results in all experiments with a confident value of 95%. Table 12 shows the comparison between neuro-fuzzy (with different compression rates) and decision tree models.

All mean f1-score values for the proposed model and random forest have a significant difference (RF had substantially better general performance than the proposed model).

It is important to note that the fuzzy granule model has considerably fewer rule elements even

Table 3. Proposed model results with different alpha values applied in the sonar dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.62	94.21	82.22
		σ	0.04		
	0.40	mean	0.72	86.61	72.41
		σ	0.06		
	0.60	mean	0.78	74.54	47.13
		σ	0.07		
	0.80	mean	0.75	54.17	21.59
		σ	0.05		
	0.95	mean	0.76	22.45	1.14
		σ	0.08		
	1.00	mean	0.76	0.00	0.00
		σ	0.07		
DT		mean	0.75		
		σ	0.02		
RF		mean	0.84		
		σ	0.02		

Table 4. Proposed model results with different alpha values applied in the wdbc dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.90	95.33	85.26
		σ	0.06		
	0.40	mean	0.94	87.56	73.03
		σ	0.02		
	0.60	mean	0.96	77.53	47.73
		σ	0.01		
	0.80	mean	0.95	60.14	30.00
		σ	0.02		
	0.95	mean	0.96	33.50	10.00
		σ	0.02		
	1.00	mean	0.96	0.00	0.00
		σ	0.01		
DT		mean	0.92		
		σ	0.00		
RF		mean	0.96		
		σ	0.00		

Table 5. Proposed model results with different alpha values applied in the credit-g dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.66	95.17	84.78
		σ	0.02		
	0.40	mean	0.63	90.59	71.51
		σ	0.03		
	0.60	mean	0.59	90.85	72.71
		σ	0.02		
	0.80	mean	0.63	90.88	73.57
		σ	0.02		
	0.95	mean	0.62	90.81	73.25
		σ	0.02		
	1.00	mean	0.65	90.82	71.95
		σ	0.03		
DT		mean	0.62		
		σ	0.02		
RF		mean	0.67		
		σ	0.01		

Table 6. Proposed model results with different alpha values applied in the iris dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.95	83.15	61.54
		σ	0.03		
	0.40	mean	0.93	70.51	51.43
		σ	0.03		
	0.60	mean	0.92	55.84	37.14
		σ	0.11		
	0.80	mean	0.95	36.71	8.57
		σ	0.04		
	0.95	mean	0.96	17.33	0.00
		σ	0.03		
	1.00	mean	0.95	0.00	0.00
		σ	0.04		
DT		mean	0.95		
		σ	0.00		
RF		mean	0.96		
		σ	0.00		

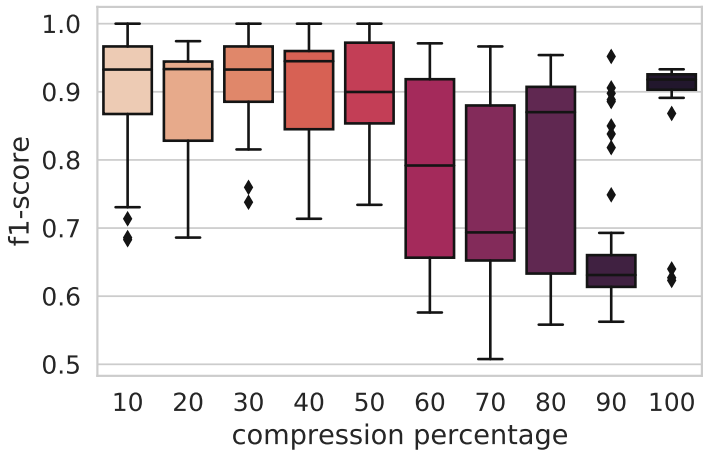

Fig. 5. The relationship between performance and model compression percentage in terms of antecedents

Table 7. Proposed model results with different alpha values applied in the wine dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.73	89.40	67.39
		σ	0.18		
	0.40	mean	0.92	76.86	58.97
		σ	0.07		
	0.60	mean	0.89	67.97	53.66
		σ	0.18		
	0.80	mean	0.95	46.09	19.51
		σ	0.04		
	0.95	mean	0.96	16.92	2.44
		σ	0.02		
	1.00	mean	0.93	0.00	0.00
		σ	0.07		
DT		mean	0.94		
		σ	0.01		
RF		mean	0.97		
		σ	0.01		

in configurations with no significant difference in the mean f1-scores with respect to the DT model.

According to the validation, there is a significant difference between models where alpha is 0.2, 0.95, and 1; in the first value, which the model compresses the most the rules (around 82%), the initial decision tree model result is higher by 0.02. However, in alpha values 0.95 and 1, which the model comprises the lowest the rules (around 41% and 39%, respectively), the proposed model is higher by 0.03 and 0.02, respectively. There is no significant difference between model results in intermediate alpha values, although the decision sets were compressed between 71% and 49%.

Table 8. Proposed model results with different alpha values applied in the ionosphere dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.79	94.72	80.00
		σ	0.09		
	0.40	mean	0.84	86.34	57.80
		σ	0.04		
	0.60	mean	0.85	77.94	37.39
		σ	0.08		
	0.80	mean	0.86	59.22	18.42
		σ	0.04		
	0.95	mean	0.86	39.28	3.48
		σ	0.05		
	1.00	mean	0.86	34.21	3.54
		σ	0.04		
DT		mean	0.86		
		σ	0.02		
RF		mean	0.93		
		σ	0.01		

Table 9. Proposed model results with different alpha values applied in the spambase dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.92	97.93	97.00
		σ	0.01		
	0.40	mean	0.92	96.71	95.07
		σ	0.02		
	0.60	mean	0.92	96.63	94.58
		σ	0.01		
	0.80	mean	0.91	96.56	94.71
		σ	0.01		
	0.95	mean	0.92	96.60	94.99
		σ	0.01		
	1.00	mean	0.92	96.56	94.71
		σ	0.02		
DT		mean	0.84		
		σ	0.01		
RF		mean	0.94		
		σ	0.00		

Table 10. Proposed model results with different alpha values applied in the diabetes dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.65	95.45	87.90
		σ	0.05		
	0.40	mean	0.68	88.11	71.95
		σ	0.02		
	0.60	mean	0.66	82.19	61.26
		σ	0.02		
	0.80	mean	0.66	81.86	60.84
		σ	0.02		
	0.95	mean	0.69	81.91	62.10
		σ	0.04		
	1.00	mean	0.63	81.82	59.85
		σ	0.04		
DT		mean	0.66		
		σ	0.01		
RF		mean	0.73		
		σ	0.01		

Table 11. Proposed model results with different alpha values applied in the abalone dataset

model	alpha		f1-score	ER	RR
F-GrM	0.20	mean	0.62	96.85	88.57
		σ	0.02		
	0.40	mean	0.62	96.84	87.89
		σ	0.01		
	0.60	mean	0.62	96.84	88.37
		σ	0.02		
	0.80	mean	0.61	96.82	88.50
		σ	0.01		
	0.95	mean	0.62	96.85	88.51
		σ	0.02		
	1.00	mean	0.63	96.85	88.71
		σ	0.01		
DT		mean	0.56		
		σ	0.00		
RF		mean	0.64		
		σ	0.00		

8 Conclusion and Future Work

This work addressed the problem of input space partition to create the base rule structure for a FIS by selecting the meaningful elements from a decision tree model given a compression rate value (controlled by the hyperparameter α).

The motivation for choosing a subset of the elements in rule-based systems is to maintain their structure as simple as possible, directly impacting their interpretability. Fewer antecedents lead to understanding the phenomena with less effort; also, the number of rules affects too. In order to build a knowledge base with the greatest interpretability, it is necessary to create it with fewer antecedents and rules.

The antecedents were characterized as type-one granules by ranges, while the implications were represented as type-two granules through relationships between granules of the lower level. The abstraction of decision sets is made by hierarchical structure between the different levels in the granules. The transition from one level to another is defined through the granulation and degranulation operations.

The methodology for extracting the relevant elements was based on the principle of justifiable granularity, which has a higher value of specificity and coverage. In this approach, the optimization method selects all level granules simultaneously through the hierarchy established.

The reconstruction of the selected granules creates a new decision set with fewer elements than the original one due to the optimization restrictions: i) minimize the number of level-one granules; ii) the cumulative Pareto front values must be equal or less than a regularizer hyper-parameter α; iii) a hyperparameter ξ restricts the number of level-one granules to consider in the selection.

The presented work defined a hierarchical measurement that helps to consider the best lower-level information granules better suited to form high-quality higher-level Information Granules. This well-formed Information Granules search is defined as an optimization problem with restrictions. The objective is to find the smallest set of best lower-level granules that maximize the hierarchical quality measurement composed by the Pareto front at different levels of specificity and coverage. As the number of elements decreases in the rule-based system, the bias increases; to increase the model variance, the resulting pruned decision set is converted to a Sugeno-type neuro-fuzzy model for classification tasks that is further optimized. The proposed Sugeno-type neuro-fuzzy model for classification has the following characteristics:

— The fuzzification layer is not fully connected, which prevents incoherent rule formation. For the sake of interpretability, the rules must be coherent and sound to the data scientist.

— The implications layer is also not fully connected, which reduces ambiguity in the output, maintaining separated rule contribution for each class; this characteristic tends to analyze the conditions to belong to a given class more easily.

— Sigmoid activation functions transform the output of the implication layer to get interpretable outcomes for classification tasks.

— In order to increase the output interpretability, is used a softmax layer to get the outcome of the model in a probability fashion that helps to suit the belongingness to the target classes better.

The results support that using a FIS with the proposed method for rule selection can compress the contained information in a classical decision set without significantly compromising the performance model. The different compression rate values (α) impact the model performance.

As shown in the results, a higher compression rate (lower α value, e.g., 0.2) degrades the model performance significantly; however, this small decision set in a complex domain might be helpful to have a more interpretable model to understand phenomena better.

An attractive characteristic is that intermediate compressing rate values (e.g., between 0.4 and 0.8) achieve a considerable reduction in the decision set without significant difference performance.

Table 12. Paired sample t-test of mean f1-scores for formal comparison between the decision tree model and the resulting fuzzy rule-based model constructed by the proposed methodology

	alpha	**0.20**	**0.40**	**0.60**	**0.80**	**0.95**	**1.00**
DT	mean	**0.79***	0.80	0.79	0.80	0.80	0.80
	σ	0.13	0.15	0.14	0.14	0.14	0.14
Fuzzy-GrM	mean	0.77	**0.81**	**0.81**	**0.82**	**0.83***	**0.82***
	σ	0.14	0.13	0.15	0.14	0.14	0.14
P-value		0.003	0.069	0.102	0.126	5.68×10^{-4}	4.32×10^{-2}
T-student		2.86	1.50	1.28	-1.15	-3.45	-1.74
DF		49	49	49	49	49	49
Significant difference		YES	NO	NO	NO	YES	YES
Element reduction	mean	93.91%	87.21%	81.02%	71.20%	58.61%	49.09%
Rule reduction	mean	82.51%	71.61%	61.38%	49.09%	41.10%	39.48%

* shows significant difference.

Table 13. Proposed model results with different alpha values applied in the creditcard dataset

model	**alpha**		**f1-score**	**ER**	**RR**
F-GrM	0.20	mean	0.90	97.63	92.20
		σ	0.02		
	0.40	mean	0.89	94.40	79.70
		σ	0.02		
	0.60	mean	0.90	92.23	75.89
		σ	0.02		
	0.80	mean	0.90	91.86	76.02
		σ	0.02		
	0.95	mean	0.90	91.81	75.77
		σ	0.02		
	1.00	mean	0.90	91.79	76.53
		σ	0.02		
DT		mean	0.88		
		σ	0.00		
RF		mean	0.93		
		σ	0.00		

In problems where the performance is crucial, higher α values are recommended, which less compress the decision set but still might be considerable (e.g., at least a mean of 39% in this experimental setup). The rule compression achieved by the proposed method is relevant in the context of IML due that simplifies the decision model by i) reducing the number of elements (antecedents and rules), decreases as well the complexity of the systems, which improves the interpretability; ii) fuzzy logic brings an interface in natural language to the human and promotes

a better understanding of the model; iii) the compression of the model can be controlled by α hyper-parameter and be set to the most convenient value for a specific application domain (see figure 3 compare the behavior of α in different domain applications).

This proposal opens the opportunity to further research in decision-set-based Information Granules for smaller and more interpretable models; this methodology can be used with different models/methods that generate decision sets. An extension of the current work is considered to design a higher-level information granule, specificity and coverage metric to select the most relevant decision set source elements for ensemble methods.

Another possible extension of this work is to use type-2 fuzzy logic, which better manages decision processes under uncertainty and achieves better performance with a smaller knowledge base in terms of rules. The overlapping of rule partition with different consequent could characterize uncertainty in higher-level fuzzy sets.

The proposed Hierarchical Decision Granules Optimization method can be adapted to any rule-based system by defining specificity and coverage metrics for each granule level. It can be interesting to incorporate different information frameworks such as probabilistic and rough sets to enhance the intrinsic semantic meaning in the Information Granule, therefore generate richer explanations taking advantage of the natural language interface that brings the fuzzy logic.

Acknowledgments

This research was supported by CONACyT (Consejo Nacional de Ciencia y Tecnología) with grant number 691247.

References

1. **Adhyaru, D. M., Patel, J., Gianchandani, R. (2010).** Adaptive neuro-fuzzy inference system based control of robotic manipulators. ICMET 2010 - 2010 International Conference on Mechanical and Electrical Technology, Proceedings, pp. 353–358.
2. **An, S., Hu, Q., Wang, C. (2021).** Probability granular distance-based fuzzy rough set model. Applied Soft Computing, Vol. 102.
3. **Azadeh, A., Gaeini, Z., Motevali Haghighi, S., Nasirian, B. (2016).** A unique adaptive neuro fuzzy inference system for optimum decision making process in a natural gas transmission unit. Journal of Natural Gas Science and Engineering, Vol. 34, pp. 472–485.
4. **Bandyopadhyay, S., Yao, J., Zhang, Y. (2017).** Granular computing with compatibility based intuitionistic fuzzy rough sets. Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, Institute of Electrical and Electronics Engineers Inc., pp. 378–383.
5. **Bargiela, A., Pedrycz, W. (2003).** Granular Computing.
6. **Bello, M., Nápoles, G., Vanhoof, K., Bello, R. (2021).** Data quality measures based on granular computing for multi-label classification. Information Sciences, Vol. 560, pp. 51–67.
7. **Brasoveanu, A., Moodie, M., Agrawal, R. (2020).** Textual evidence for the perfunctoriness of independent medical reviews. CEUR Workshop Proceedings, Vol. 2657, pp. 1–9.
8. **Castillo, O., Cervantes, L., Soria, J., Sanchez, M., Castro, J. R. (2016).** A generalized type-2 fuzzy granular approach with applications to aerospace. Information Sciences, Vol. 354, pp. 165–177.
9. **Chan, V. K. H., Chan, C. W. (2020).** Towards explicit representation of an artificial neural network model: Comparison of two artificial neural network rule extraction approaches. Petroleum, Vol. 6, No. 4, pp. 329–339.
10. **Chen, Y., Miao, D. (2020).** Granular regression with a gradient descent method. Information Sciences, Vol. 537, pp. 246–260.
11. **Chen, Y., Zhu, S., Li, W., Qin, N. (2021).** Fuzzy granular convolutional classifiers. Fuzzy Sets and Systems.
12. **Cheng, M. Y., Tsai, H. C., Ko, C. H., Chang, W. T. (2008).** Evolutionary fuzzy neural inference system for decision making in geotechnical engineering. Journal of Computing in Civil Engineering, Vol. 22, No. 4, pp. 272–280.
13. **Cui, H., Yue, G., Zou, L., Liu, X., Deng, A. (2021).** Multiple multidimensional linguistic reasoning algorithm based on property-oriented

linguistic concept lattice. International Journal of Approximate Reasoning, Vol. 131, pp. 80–92.

14. **De Medeiros, I. B., Soares-Machado, M. A., Damasceno, W. J., Caldeira, A. M., Dos-Santos, R. C., Da-Silva Filho, J. B. (2017).** A Fuzzy Inference System to Support Medical Diagnosis in Real Time. Vol. 122, pp. 167–173.
15. **Deshpande, S. U., Bhosale, S. S. (2013).** Adaptive neuro-fuzzy inference system based robotic navigation. 2013 IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC, IEEE Computer Society.
16. **Ding, W., Wang, J., Wang, J. (2020).** Multi-granulation consensus fuzzy-rough based attribute reduction. Knowledge-Based Systems, Vol. 198.
17. **Ding, X., Zeng, Z., Lun, L. (2010).** Granular neural networks computing on fuzzy information table. Vol. 1, pp. 412–418.
18. **Gaeta, A., Loia, V., Orciuoli, F. (2021).** A comprehensive model and computational methods to improve Situation Awareness in Intelligence scenarios. Applied Intelligence, Vol. 51, No. 9, pp. 6585–6608.
19. **Gayathri, B. M., Sumathi, C. P. (2016).** Mamdani fuzzy inference system for breast cancer risk detection. **Karthikeyan M., K. N.**, editor, 2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2015, Institute of Electrical and Electronics Engineers Inc.
20. **Ghiasi, B., Sheikhian, H., Zeynolabedin, A., Niksokhan, M. H. (2020).** Granular computing-neural network model for prediction of longitudinal dispersion coefficients in rivers. Water Science and Technology, Vol. 80, No. 10, pp. 1880–1892.
21. **Guan, Q., Guan, J. H. (2014).** Knowledge acquisition of interval set-valued based on granular computing. Applied Mechanics and Materials, Vol. 543-547, pp. 2017–2023.
22. **Guo, H., Wang, W. (2019).** Granular support vector machine: a review. Artificial Intelligence Review, Vol. 51, No. 1, pp. 19–32.
23. **Guo, S., Zhao, H. (2021).** Hierarchical classification with multi-path selection based on granular computing. Artificial Intelligence Review, Vol. 54, No. 3, pp. 2067–2089.
24. **Honka, A. M., Van Gils, M. J., Pärkkä, J. (2011).** A personalized approach for predicting the effect of aerobic exercise on blood pressure using a Fuzzy Inference System. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 8299–8302.
25. **Hu, X., Pedrycz, W., Wu, K., Shen, Y. (2021).** Information granule-based classifier: A development of granular imputation of missing data. Knowledge-Based Systems, Vol. 214.
26. **Jiang, F., Chen, Y. M. (2015).** Outlier detection based on granular computing and rough set theory. Applied Intelligence, Vol. 42, No. 2, pp. 303–322.
27. **Leite, D., Costa, P., Gomide, F. (2013).** Evolving granular neural networks from fuzzy data streams. Neural Networks, Vol. 38, pp. 1–16.
28. **Li, D., Miao, D., Du, W. (2006).** Application of granular computing to artificial neural network. Tongji Daxue Xuebao/Journal of Tongji University, Vol. 34, No. 7, pp. 960–964.
29. **Li, W., Luo, Y., Tang, C., Zhang, K., Ma, X. (2021).** Boosted Fuzzy Granular Regression Trees. Mathematical Problems in Engineering, Vol. 2021, pp. 9958427.
30. **Li, W., Ma, X., Chen, Y., Dai, B., Chen, R., Tang, C., Luo, Y., Zhang, K. (2021).** Random Fuzzy Granular Decision Tree. Mathematical Problems in Engineering, Vol. 2021, pp. 557–682.
31. **Liu, H., Gegov, A., Cocea, M. (2016).** Rule-based systems: a granular computing perspective. Granular Computing, Vol. 1, No. 4, pp. 259–274.
32. **Mencar, C., Fanelli, A. M. (2008).** Interpretability constraints for fuzzy information granulation. Information Sciences, Vol. 178, No. 24, pp. 4585–4618.
33. **Moreno, J. E., Sanchez, M. A., Mendoza, O., Rodríguez-Díaz, A., Castillo, O., Melin, P., Castro, J. R. (2020).** Design of an interval Type-2 fuzzy model with justifiable uncertainty. Information Sciences, Vol. 513, pp. 206–221.
34. **Panoutsos, G., Mahfouf, M. (2007).** Information fusion using Granular Computing Neural-Fuzzy Networks and expert knowledge. 2007 European Control Conference, ECC 2007, Institute of Electrical and Electronics Engineers Inc., pp. 776–782.
35. **Panoutsos, G., Mahfouf, M., Mills, G. H., Brown, B. H. (2010).** A generic framework for enhancing the interpretability of granular computing-based information. 2010 IEEE International Conference on Intelligent Systems, IS 2010 - Proceedings, pp. 19–24.

36. **Pedrycz, A., Hirota, K., Pedrycz, W., Dong, F. (2012).** Granular representation and granular computing with fuzzy sets. Fuzzy Sets and Systems, Vol. 203, pp. 17–32.
37. **Pedrycz, W. (2010).** Human centricity in computing with fuzzy sets: An interpretability quest for higher order granular constructs. Journal of Ambient Intelligence and Humanized Computing, Vol. 1, No. 1, pp. 65–74.
38. **Pedrycz, W., Homenda, W. (2013).** Building the fundamentals of granular computing: A principle of justifiable granularity. Applied Soft Computing Journal, Vol. 13, No. 10, pp. 4209–4218.
39. **Pedrycz, W., Vukovich, G. (2001).** Granular neural networks. Neurocomputing, Vol. 36, No. 1-4, pp. 205–224.
40. **Peters, G., Lacic, Z. (2012).** Tackling outliers in granular box regression. Information Sciences, Vol. 212, pp. 44–56.
41. **Pota, M., Esposito, M., De Pietro, G. (2017).** Designing rule-based fuzzy systems for classification in medicine. Knowledge-Based Systems, Vol. 124, pp. 105–132.
42. **Qian, Y., Zhang, H., Sang, Y., Liang, J. (2014).** Multigranulation decision-theoretic rough sets. International Journal of Approximate Reasoning, Vol. 55, No. 1 part 2, pp. 225–237.
43. **Sanchez, M. A., Castillo, O., Castro, J. R., Melin, P. (2014).** Fuzzy granular gravitational clustering algorithm for multivariate data. Information Sciences, Vol. 279, pp. 498–511.
44. **Solis, A. R., Panoutsos, G. (2013).** Granular computing neural-fuzzy modelling: A neutrosophic approach. Applied Soft Computing Journal, Vol. 13, No. 9, pp. 4010–4021.
45. **Sugeno, M., Kang, G. T. (1988).** Structure identification of fuzzy model. Fuzzy Sets and Systems, Vol. 28, No. 1, pp. 15–33.
46. **Truong, H. Q., Ngo, L. T., Pham, L. T. (2019).** Interval type-2 fuzzy possibilistic c-means clustering based on granular gravitational forces and particle swarm optimization. Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 23, No. 3, pp. 592–601.
47. **Tuan, T. M., Lan, L. T. H., Chou, S.-Y., Ngan, T. T., Son, L. H., Giang, N. L., Ali, M. (2020).** M-CFIS-R: Mamdani complex fuzzy inference system with rule reduction using complex fuzzy measures in granular computing. Mathematics, Vol. 8, No. 5.
48. **Vasilev, N., Mincheva, Z., Nikolov, V. (2020).** Decision tree extraction using trained neural network. Smartgreens 2020 - Proceedings of the 9th International Conference on Smart Cities and Green ICT Systems, pp. 194–200.
49. **Wang, W., Xiong, S. (2013).** Research of logical reasoning and application based on granular computing rough sets. Advanced Materials Research, Vol. 622, pp. 1877–1881.
50. **Wu, Q., Wang, P., Huang, X., Yan, S. (2005).** Adaptive discretizer for machine learning based on granular computing and rough sets. 2005 IEEE International Conference on Granular Computing, volume 2005, pp. 292–295.
51. **Xie, K., Xie, J., Du, L., Xu, X. (2009).** Granular computing and neural network integrated algorithm applied in fault diagnosis. 6th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2009, volume 1, pp. 188–191.
52. **Xiong, C., Qian, W., Wang, Y., Huang, J. (2021).** Feature selection based on label distribution and fuzzy mutual information. Information Sciences, Vol. 574, pp. 297–319.
53. **Xu, J., Miao, D., Zhang, Y., Zhang, Z. (2017).** A three-way decisions model with probabilistic rough sets for stream computing. International Journal of Approximate Reasoning, Vol. 88, pp. 1–22.
54. **Xu, X., Wang, G., Ding, S., Jiang, X., Zhao, Z. (2015).** A new method for constructing granular neural networks based on rule extraction and extreme learning machine. Pattern Recognition Letters, Vol. 67, pp. 138–144.
55. **Xue, M., Duan, X., Liu, W., Ren, Y. (2020).** A semantic facial expression intensity descriptor based on information granules. Information Sciences, Vol. 528, pp. 113–132.
56. **Yan, E., Song, J., Ren, Y., Zheng, C., Mi, B., Hong, W. (2020).** Construction of three-way attribute partial order structure via cognitive science and granular computing. Knowledge-Based Systems, Vol. 197.
57. **Yang, J. G., Kim, J. K., Kang, U. G., Lee, Y. H. (2014).** Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (ANFIS-LDA). Personal and Ubiquitous Computing, Vol. 18, No. 6, pp. 1351–1362.
58. **Zadeh, L. A. (1965).** Fuzzy sets. Information and Control, Vol. 8, No. 3, pp. 338–353.

59. **Zadeh, L. A. (1975).** The concept of a linguistic variable and its application to approximate reasoning–I. Information Sciences, Vol. 8, No. 3, pp. 199–249.

60. **Zein-Sabatto, S., Mikhail, M., Bodruzzaman, M., DeSimio, M., Derriso, M. (2013).** Multistage fuzzy inference system for decision making and fusion in fatigue crack detection of aircraft structures. AIAA Infotech at Aerospace (I at A) Conference.

61. **Zhang, X., Miao, D. (2014).** Quantitative information architecture, granular computing and rough set models in the double-quantitative approximation space of precision and grade. Information Sciences, Vol. 268, pp. 147–168.

62. **Zhang, Y., Zhu, X., Huang, Z. (2009).** Fuzzy sets based granular logics for granular computing. Proceedings - 2009 International Conference on Computational Intelligence and Software Engineering, CiSE 2009.

63. **Zhang, Z.-J., Huang, J., Wei, Y. (2015).** FI-FG: Frequent item sets mining from datasets with high number of transactions by granular computing and fuzzy set theory. Mathematical Problems in Engineering, Vol. 2015.

64. **Zhou, D., Dai, X. (2015).** Combining granular computing and RBF neural network for process planning of part features. International Journal of Advanced Manufacturing Technology, Vol. 81, No. 9-12, pp. 1447–1462.

65. **Zhou, J., Lai, Z., Gao, C., Miao, D., Yue, X. (2018).** Rough possibilistic C-means clustering based on multigranulation approximation regions and shadowed sets. Knowledge-Based Systems, Vol. 160, pp. 144–166.

66. **Zhou, P., Xu, Z., Zhao, J., Song, C., Shao, Z. (2021).** Long-term hybrid prediction method based on multiscale decomposition and granular computing for oxygen supply network. Computers and Chemical Engineering, Vol. 153.

67. **Zhu, X., Pedrycz, W., Li, Z. (2018).** Granular representation of data: A design of families of epsilon-Information granules. IEEE Transactions on Fuzzy Systems, Vol. 26, No. 4, pp. 2107–2119.

68. **Zhu, X., Pedrycz, W., Li, Z. (2021).** A Development of Granular Input Space in System Modeling. IEEE Transactions on Cybernetics, Vol. 51, No. 3, pp. 1639–1650.

69. **Ziqi, Z., Xinting, T., Xiaofeng, Z., Hongjiang, G., Kun, L. (2017).** Research of Rough Set Model under Logical Computing of Granular. Proceedings - 2017 IEEE International Conference on Computational Science and Engineering and IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, CSE and EUC 2017, volume 1, Institute of Electrical and Electronics Engineers Inc., pp. 333–336.

Article received on 29/06/2021; accepted on 17/11/2021.
Corresponding author is Raúl Navarro-Almanza.