Int. J. Fuzzy Syst. (2021) 23(5):1431–1452
https://doi.org/10.1007/s40815-020-00968-w

Continuous Linguistic Variables and Their Applications to Data Mining and Time Series Prediction

Erick González-Caballero[1] · Rafael A. Espín-Andrade[2] · Witold Pedrycz[3,4,5] · Luis Martínez[6] · Liliana A. Guerrero-Ramos[7]

Received: 11 January 2020 / Revised: 15 September 2020 / Accepted: 18 September 2020 / Published online: 5 February 2021

Abstract Membership function estimation is one of the less explored, albeit important, areas in fuzzy sets. This paper aims to define a new family of fuzzy sets called *general continuous linguistic variables* (GCLV), which represents a linguistic variable rather than a set of linguistic values. We refer to it as the *principle of representation of linguistic variables*. They are based on the well-known sigmoidal functions and contain at least three different classes of membership functions, namely, an increasing sigmoidal function, a decreasing sigmoidal function, and a convex one. These diverse features are essential to represent linguistic values exhibiting different semantics. We explore the properties of GCLV, including those ones over that allow us to approximate every continuous membership function. Finally, we illustrate the applicability of GCLV as a fuzzy tool. This leads to the development of the foundations of a new vehicle in fuzzy sets useful in data mining and time series prediction.

Keywords Membership function · Sigmoidal function · Linguistic variable · Data mining · Time series prediction

✉ Erick González-Caballero
erickgc@yandex.com

Rafael A. Espín-Andrade
rafaelalejandroespinandrade@gmail.com

Witold Pedrycz
wpedrycz@ualberta.ca

Luis Martínez
martin@ujaen.es

Liliana A. Guerrero-Ramos
liliguer@yahoo.com

1 Technical University of Havana, 27 Calle 114 # 11901 e/ Ciclovía y Rotonda, Marianao, Havana, Cuba

2 Autonomous University of Coahuila, Blvd Revolución No 151 Oriente, Torreón, Coahuila, Mexico

3 University of Alberta Edmonton, Edmonton, AB T6R 2V4, Canada

4 King Abdulaziz University Jeddah, 21589 Jeddah, Saudi Arabia

5 Systems Research Institute, Polish Academy of Sciences Warsaw, Warsaw, Poland

6 University of Jaén, 23071 Jaén, Spain

7 Autonomous University of Coahuila, Blvd Revolución No 151 Oriente, Torreón, Coahuila, Mexico

1 Introduction

Historically, studies on membership functions have been limited in comparison with other topics studied in fuzzy sets. Even though there exists a wide range of descriptions of many of these functions, very often the design of fuzzy tools does not consider what kind of membership function is the most suitable for a given application.

The first approach comes with the idea of a fuzzy set [1]. Dombi paid attention to this problem in [2, 3], in which common requirements concerning membership functions were summarized. Some of them are listed below:

1. Membership functions are continuous.
2. All membership functions are either monotonically increasing, or monotonically decreasing or can be divided into parts where they are monotonically increasing or decreasing.
3. The monotonous functions are either convex or concave or there exists a point of inflexion which divides the function into a convex and concave parts; they are s-shaped functions or z-shaped functions.

General criteria to construct membership functions included the following views [4, 5]: (1) likelihood view, (2) random set view, (3) similarity view, (4) utility view and (5) measurement view.

Usually, membership functions used in fuzzy systems, like Fuzzy Inference Systems, are the simplest ones such as trapezoidal and triangular; see [6, 7]. Other approaches adjust the data by interpolation methods [8] or by linear functions defined over some subintervals [9].

In [10], Valente de Oliveira advocated that interpretable membership functions should satisfy the following requirements:

1. The number of membership functions should be 7 ± 2, [11]. There exists some psychological justification behind this: 7 ± 2 is the number of entities that people process in short-term memory.
2. Distinguishability. Every membership function should represent a linguistic value with a clear semantics.
3. Normality. Membership functions should be normal.
4. Natural Zero Positioning. One of the membership functions should represent the value "nearly zero" for this end it is recommended to be unimodal, convex and centered at zero.
5. Coverage. Every piece of data should have linguistic representation.

Valente de Oliveira work's purpose is somewhat similar to the objective of this study. He notes that usually fuzzy systems are concerned almost exclusively about accurate results. However, this focal trend contradicts the essence of fuzzy sets to obtain semantically sound (justified) results, which are interpretable in the form of linguistic terms. He proposes semantic constraints to optimize membership functions. These constraints are the ones previously mentioned. In further investigations, we revisit these ideas.

Drakopoulos [12], developed a theory of sigmoidal membership functions and his Sigmoidal Bubble Theorem forms the basis to approximate every membership function by sigmoidals. This method is used to approximate membership functions by piece-wise sigmoidal functions. This approximation is limited with regard to the semantic of the predicates, because it is difficult to interpret a compound predicate based on the simple ones. It was applied to pattern recognition [13].

Sigmoidal functions entail an assumption that the change of the belief degree that "x is A" is proportional to the belief degree that "x is A" and the belief degree that "x is not A" [5]. It is a special type of similarity to construct a membership function because it is a sort of distance between some value and a desirable value γ. It has been widely used in artificial neural networks regarded as universal approximators [14].

This paper introduces a new parametrized family of membership functions, called *general continuous linguistic variables* (GCLV) based on four parameters, whose aim is to adjust each GCLV from experimental data by optimizing the truth value of compound predicates, such that the GCLV are atoms with a semantic meaning.

Its advantages over other parametrized families are outlined as follows:

1. It contains functions of at least three kind of shapes, one is an strictly increasing membership function, other is an strictly decreasing and the third is a convex one which is strictly increasing in its first part and strictly decreasing in its second part. These different types of shape allow to represent different linguistic values. Usually the families of membership functions used in literature are of the same shape and lack of expressiveness [15].
2. The members of the GCLVs are modified by linguistic hedges. This property increases the expressiveness of the results.
3. The GCLV incorporates the possibilities like universal approximator relevant in the setting sigmoidal functions. Every continuous membership function can be approximated by the members of this family.
4. Its parameters have a meaning as presented in Dombi's approach.
5. It is possible that GCLVs satisfy the conditions suggested by Valente de Oliveira in [10] to guarantee the semantics.

We formulate the concept of a so-called *principle of representation of linguistic variables*, which means that part of the family of GCLVs represents a linguistic variable, like, e.g., "age" or "height".

Each linguistic value is associated with a fuzzy set by an specific 4-tuple of parameters, because of the different shapes, it is possible to identify a single family with a linguistic variable, in which the most important linguistic values can be represented by fuzzy sets, only fixing the values of four parameters. Therefore, the aim of adjusting experimental data optimizing over the space of parameters can be achieved.

Our motivation is to develop the foundations of a new tool in fuzzy theory, which is useful in Data Mining. The main novelty is that data can be adjusted from a data set where the output is a linguistic value with many possible different semantics. It is a type of linguistic mining completed by the optimization on the space of parameters.

The paper is organized as follows. Section 2 summarizes the main concepts necessaries to understand this paper. Section 3 contains the main definitions of the paper. Section 4 explores the parametric meanings and properties of the family according to Dombi's theory. Section 5 elaborates on a semantic approach, where algorithms are designed to obtain linguistic interpretations of the membership functions.

Section 6 is dedicated to illustrate the possible applications of this new family in fuzzy tools. The paper is concluded in Sect. 7.

2 Basic Concepts

It is well-known that the sigmoidal membership function with parameters $\alpha > 0$ and $\gamma \in \mathbb{R}$, is defined by the following expression:

$$sigm(x; \alpha, \gamma) = \frac{1}{1 + e^{-\alpha(x-\gamma)}}. \qquad (1)$$

It is a solution to the differential equation:

$$\frac{dX}{\mathrm{d}t} = \alpha X(1 - X). \qquad (2)$$

In other words, it is considered that the marginal increase of the belief degree that "x is A" is proportional to the belief degree that "x is A" and the belief degree that "x is not A" [5].

It can be easily proved the following property being extensively used in this paper:

$$1 - sigm(x; \alpha, \gamma) = sigm(x; -\alpha, \gamma).$$

Dombi's work [2, 3] is dedicated to the membership functions, in which a list of objectives behind the formation of membership functions is identified:

1. On a theoretical basis,
2. Easy to calculate and fit to the problem,
3. Described by only a few parameters,
4. With parameters that are meaningful,
5. With a linearized form for the applications, and
6. With membership and operators closely connected.

Finally, four parameters are fixed to define a membership function, two for the interval (a, b) where the function is defined, λ meaning the *sharpness* and v the *decision level*, i.e., the value which is mapped by the membership function to 0.5.

A membership function which satisfies the previous conditions and additionally which contains properties of negation, conjunction and disjunction operators was built.

The definition and a detailed study of t-norms are covered in [16]. T-norms offer an axiomatic formalization to model conjunction. These axioms are commutativity, associativity, monotonicity and a boundary condition where 1 is the neutral element.

3 General Continuous Linguistic Variables

In this section, we introduce a new kind of parametric membership functions with the characteristic that they can take many shapes. The rationale of this approach is that each shape, e.g., triangular, trapezoidal, Gaussian or sigmoid, can be associated with only single semantic. Moreover, less accurate fuzzy systems are modeled with the usage. Otherwise, many-shape membership functions can be translated to many semantics provided of higher accuracy. The accuracy is the consequence of the flexibility of those functions, which can adapt its shape to the data. The formal definition is given in the following.

Definition 1 A general continuous linguistic variable (GCLV) is defined as:

$$\mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0)$$
$$= T(sigm^m(x; \alpha, \gamma), (1 - sigm(x; \alpha, \gamma))^{m_0 - m}),$$

where $m \in [0, m_0]$, $m_0 > 0$ is fixed, $\alpha > 0$, T is a t-norm and $sigm(x; \alpha, \gamma)$ is a sigmoidal membership function with parameters α and $\gamma \in \mathbb{R}$.

Remark 1 Here, we consider $0^0 = 1$.

See that the maximum of the GCLV can be smaller than 1, therefore, below another membership function is defined which allows to change the range of the GCLV.

Definition 2 A scaled general continuous linguistic variable (SGCLV) is defined by:

$$\mathrm{SGCLV}_{C,T}(x; \alpha, \gamma, m, m_0) = C \cdot \mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0),$$

where $C > 0$ is a scalar, constrained by the condition $\mathrm{SGCLV}_{C,T}(x; \alpha, \gamma, m, m_0) \leq 1$.

Further, we define a kind of SGCLV function representing normal fuzzy sets. This kind of membership functions are important for interpretability.

Definition 3 A normalized general continuous linguistic variable (NGCLV) is defined in the following form:

$$\mathrm{NGCLV}_T(x; \alpha, \gamma, m, m_0) = \frac{\mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0)}{M},$$

where M is the maximum of the GCLV, if it exists.

Remark 2 $\lim_{x \to +\infty} sigm(x; \alpha, \gamma) = 1$ and $\lim_{x \to -\infty} sigm(x; \alpha, \gamma) = 0$, hence, for $m \in (0, m_0)$, $\lim_{x \to -\infty} \mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0) = \lim_{x \to +\infty} \mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0) = 0$. $m = m_0$ or $m = 0$ represent the sigmoidal and the NOT sigmoidal, respectively, and $M = 1$.

Proposition 1 *The GCLVs have always an upper bound in* $\mathbb{R}$.

Proof First, let us consider $m \in (0, m_0)$ and some $\epsilon > 0$. There exists a compact interval [*a*, *b*] such that

$\mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0) \geq \epsilon$, taking into account the remark above.

$\mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0)$ has a maximum in [*a*, *b*]. This follows from the continuity of the powered sigmoidals and also the non-decreasing property of t-norms.

This property is evident when $m = 0, m_0$. □

Proposition 1 demonstrates that always exists M in Definition 3 and the normalized membership functions can be defined. Specially, in Proposition 2 we shall demonstrate and give the explicit formulas of M and $x_{\max}$, when T is the product t-norm. These formulas allow efficiency and accuracy in the application of NGCLVs in data mining.

Proposition 2 *The* $\mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0)$ *based on the product t-norm has a maximum equals to* $M = \left(\frac{m}{m_0}\right)^m \left(1 - \frac{m}{m_0}\right)^{m_0 - m}$ *in* $x_{\max} = \frac{1}{\alpha} \ln\left(\frac{m}{m_0 - m}\right) + \gamma$, *where* $m \neq 0, m_0$.

Proof Let us recall $X = sigm(x; \alpha, \gamma)$ is the solution to the differential equation $\frac{dX}{\mathrm{d}t} = \alpha X(1 - X)$.

$$\begin{aligned}
&\frac{d}{\mathrm{d}t}\left(X^m (1 - X)^{m_0 - m}\right) \\
&\quad = m X^{m-1} \frac{dX}{\mathrm{d}t} (1 - X)^{m_0 - m} \\
&\qquad + X^m (m_0 - m)(1 - X)^{m_0 - m - 1} \left(-\frac{dX}{\mathrm{d}t}\right).
\end{aligned}$$

Substituting $\frac{dX}{\mathrm{d}t}$ by $\alpha X(1 - X)$ in the equation and grouping some terms we have,

$$\begin{aligned}
&\alpha\Big(m X^m (1 - X)^{m_0 - m + 1} \\
&\quad -(m_0 - m) X^{m+1} (1 - X)^{m_0 - m}\Big) \\
&\quad = \alpha X^m (1 - X)^{m_0 - m} [m(1 - X) - (m_0 - m)X] \\
&\quad = \alpha X^m (1 - X)^{m_0 - m} (m - m_0 X).
\end{aligned}$$

Therefore, $\frac{d}{\mathrm{d}t}\left(X^m (1 - X)^{m_0 - m}\right) = 0$ if and only if $X = \frac{m}{m_0}$ or $M = \left(\frac{m}{m_0}\right)^m \left(1 - \frac{m}{m_0}\right)^{m_0 - m}$. The trivial cases $X = 0$ and $X = 1$ were excluded.

Now, let us calculate the second derivative:

$$\begin{aligned}
&\alpha \frac{d}{\mathrm{d}t}\left(X^m (1 - X)^{m_0 - m} (m - m_0 X)\right) \\
&\quad = \alpha \left[\frac{d}{\mathrm{d}t}\left(X^m (1 - X)^{m_0 - m}\right)(m - m_0 X)\right. \\
&\quad \left. + X^m (1 - X)^{m - m_0} \left(-m_0 \frac{dX}{\mathrm{d}t}\right)\right].
\end{aligned}$$

Substituting $X = \frac{m}{m_0}$, taking into account $\frac{d}{\mathrm{d}t}\left(X^m (1 - X)^{m_0 - m}\right)|_{X = \frac{m}{m_0}} = 0$ and $\alpha > 0$, if $m \neq 0, m_0$ then the second derivative is negative and therefore M is a maximum.

Finally, $sigm(x; \alpha, \gamma) = \frac{1}{1 + e^{-\alpha(x - \gamma)}}$ implies $sigm(x; \alpha, \gamma) = \frac{m}{m_0}$ if $x_{\max} = \frac{1}{\alpha} \ln\left(\frac{m}{m_0 - m}\right) + \gamma$. □

$\mathrm{GCLV}_T(x; \alpha, \gamma, m, m_0)$ is a family of membership functions, which changes its shape according to the quartet of parameters.

When $m = m_0$, we have a sigmoidal membership function, and $m = 0$ corresponds to the NOT sigmoidal membership function.

Besides, for $m \in (0, m_0)$ we obtain the family of intermediate membership functions, between the sigmoidal and the NOT sigmoidal. Here, intermediate means that if $x_{\max}$ maximizes one membership function $F(x; \alpha, \gamma, m, m_0)$ of the family for $m \in (0, m_0)$, then it is finite, taking into account that the NOT sigmoidal is maximum for $-\infty$, the sigmoidal is maximum for $+\infty$ and $-\infty < x_{\max} < +\infty$.

Usually a membership function is considered equivalent to a fuzzy set representing a linguistic value, therefore, a family of membership functions can be considered a set of fuzzy sets representing a linguistic variable. In this paper, this assertion is justified by the multiple shapes we could obtain, only changing four parameters.

For example, from a linguistic variable like "height", three linguistic values can be obtained, fixing $m_0 = 1$, and using the formula $\alpha = \alpha(\beta, \gamma)$, where $\alpha(\beta, \gamma) = \frac{\ln(0.99) - \ln(0.01)}{\gamma - \beta}$ the quartets ($\beta = 130$, $\gamma = 170$, $m = m_0 = 1$), ($\beta = 130$, $\gamma = 170$, $m_0 = 1$, $m = 0.5$) and ($\beta = 130$, $\gamma = 170$, $m_0 = 1$, $m = 0$), they represent the linguistic values: "tall", "medium" and "short", respectively; see Fig. 1. We used the product t-norm and the membership functions were normalized.

Let us define $\mathcal{G}\left(\mathcal{X}; \alpha, \gamma, \left\{1, \frac{1}{2}\right\}\right) = \{\mathrm{NGCLV}_T(x; \alpha, \gamma, m, 1) | x \in \mathcal{X},.$

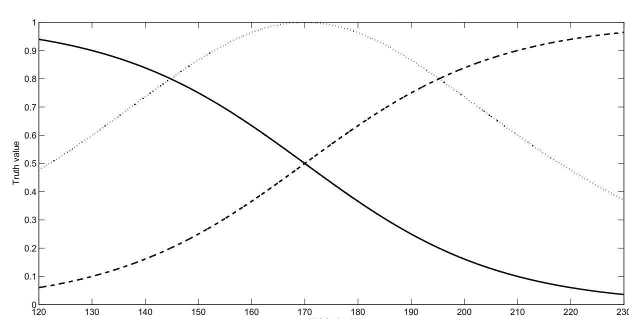


Fig. 1 Three differently shaped general continuous linguistic variables

$\alpha \in [a_\alpha, b_\alpha], \gamma \in [a_\gamma, b_\gamma], m \in \{1, \frac{1}{2}\}\}$, where $\mathcal{X}$ is a compact set, T is the product t-norm, $[a_\alpha, b_\alpha] \subseteq \mathbb{R}^*$ and $[a_\gamma, b_\gamma] \subseteq \mathbb{R}$.

$\mathcal{CMF}(\mathcal{X}) = \{f(x) | f(x)$ is a continuous membership function on $\mathcal{X}\}$.

Theorem 1 *Let us define the set $\mathcal{L}$ such that $f(x) \in \mathcal{L}$ if it satisfies one of the following conditions:*

- $f(x) \in \mathcal{G}(\mathcal{X}; \alpha, \gamma, \{1, \frac{1}{2}\})$,
- $f(x) = 1 - g(x)$, for some $g(x) \in \mathcal{G}(\mathcal{X}; \alpha, \gamma, \{1, \frac{1}{2}\})$,
- $f(x) = \max(g(x), h(x))$, *for some* $g(x), h(x) \in \mathcal{L}$,
- $f(x) = \min(g(x), h(x))$, *for some* $g(x), h(x) \in \mathcal{L}$.

Then, $\mathcal{L}$ is dense in $\mathcal{CMF}(\mathcal{X})$.

Proof Every sigmoidal membership function belongs to $\mathcal{L}$ according to the first condition of the theorem.

If the sigmoidal belongs to $\mathcal{L}$, then the NOT sigmoidal also belongs to $\mathcal{L}$, because of the second condition.

Here we considered $\alpha \in \mathbb{R}^*$ and not $\alpha > 0$, based on the second condition of the theorem and the property $1 - sigm(x; \alpha, \gamma) = sigm(x; -\alpha, \gamma)$.

Therefore, the domain of α can be extended.

Let us note that when $m = \frac{1}{2}$, the NGCLV is symmetrical in α with respect to γ, therefore, the image of one value is invariant under α and $-\alpha$. It is another justification to define $\alpha \in \mathbb{R}^*$. Further, we prove this symmetry when $m = \frac{1}{2}$.

First, suppose $x_1, x_2 \in \mathcal{X}$ and $v_1, v_2 \in]0, 1[$, $x_1 \neq x_2$ and $v_1 \neq v_2$.

There exist $\gamma \in \mathbb{R}$ and $\alpha \in \mathbb{R}^*$ such that, those x_1, x_2, v_1 and v_2 satisfy the equations $\frac{1}{1+e^{-\alpha(x_1-\gamma)}} = v_1$ and $\frac{1}{1+e^{-\alpha(x_2-\gamma)}} = v_2$.

They are, $\alpha = \frac{1}{x_1 - x_2} \ln\left(\frac{v_1(1-v_2)}{v_2(1-v_1)}\right)$ and $\gamma = \frac{1}{2}\left(x_1 + x_2 - \frac{1}{\alpha}\ln\left(\frac{v_1 v_2}{(1-v_1)(1-v_2)}\right)\right)$.

Now, suppose $v_1 = 1$ and $v_2 \in]0, 1[$. The membership function $\mathrm{NGCLV}_T(x; \alpha, x_1, \frac{1}{2}, 1)$, where $\alpha = \frac{1}{x_2 - x_1} arccosh\left(\frac{2 - v_2^2}{v_2^2}\right)$, belongs to $\mathcal{L}$.

It fulfills $\mathrm{NGCLV}_T(x_1; \alpha, x_1, \frac{1}{2}, 1) = 1$ and $\mathrm{NGCLV}_T(x_2; \alpha, x_1, \frac{1}{2}, 1) = v_2$. Given $v_1 = 0$ and $v_2 \in]0, 1[$, we have $1 - \mathrm{NGCLV}_T(x_1; \alpha, x_1, \frac{1}{2}, 1) = 0$ and $1 - \mathrm{NGCLV}_T(x_2; \alpha, x_1, \frac{1}{2}, 1) = v_2$, for $\alpha = \frac{1}{x_2 - x_1} arccosh\left(\frac{2-(1-v_2)^2}{(1-v_2)^2}\right)$. The change of v_1 by v_2 maintains the validity of the proofs.

So far, we have proved that for every $x_1, x_2 \in \mathcal{X}$ and for every $v_1, v_2 \in [0, 1]$, where $x_1 \neq x_2$, $v_1 \neq v_2$, there exists a membership function $g(x)$ of $\mathcal{L}$, such that $g(x_1) = v_1$ and $g(x_2) = v_2$, except for $v_1 = 1$ and $v_2 = 0$.

To complete this proof remains to apply the Stone Approximation Theorem, see [17], where the original range $\mathbb{R}$ is restricted to [0, 1] and we excluded functions $F(x)$, where $F(x_1) = 1$ and $F(x_2) = 0$ for some $x_1, x_2 \in \mathcal{X}$. All the hypothesis of the Stone Theorem are satisfied. In what follows, we provide the demonstration reproducing the one that appeared in [17].

There exists a function $g_{xy}(z) \in \mathcal{L}$, such that $g_{xy}(x) = F(x)$ and $g_{xy}(y) = F(y)$. Let us fix $\epsilon > 0$. F and g_{xy} are continuous, therefore, there exists an open neighborhood $U(y)$ of y, where $g_{xy}(z) > F(z) - \epsilon$ for all $z \in \mathcal{X} \cap U(y)$.

Let us fix x and select a $U(y)$ for each $y \in \mathcal{X}$. $\mathcal{X}$ is compact and hence there exists a finite set of y_is, such that $\mathcal{X} \subset \cup_{i=1}^{n} U(y_i)$. From $h_x(z) = \sup_{i=1}^{n} g_{xy_i}(z)$ it follows that $h_x(z) > F(z) - \epsilon$ and evidently $h_x(x) = F(x)$.

Besides, there exists an open neighborhood of x, $V(x)$, such that $h_x(z) < F(z) + \epsilon$. Again, we can select a finite set of x_js where $\mathcal{X} \subset \cup_{j=1}^{m} V(x_j)$.

Define $h(z) = \inf_{j=1}^{m} h_j(z)$. Evidently, $h \in \mathcal{L}$.

The two conditions $h(z) > F(z) - \epsilon$ and $h(z) < F(z) + \epsilon$ for all $z \in \mathcal{X}$ yield $|h(z) - F(z)| < \epsilon$.

It means that, every continuous membership function $MF(x)$ can be uniformly approximated by functions in $\mathcal{L}$, with the previous exceptions.

Now, suppose there exists a membership function $MF(x)$, such that for some $x_1, x_2 \in \mathcal{X}$, $MF(x_1) = 1$ and $MF(x_2) = 0$. This kind of function includes triangular and trapezoidal membership functions.

Let us fix $\epsilon > 0$ and define $\overline{MF}(x)$ such that $\overline{MF}(x) = MF(x)$, for $x \in \mathcal{X} \backslash S$, where $S = \{x \in \mathcal{X} | MF(S) = 0\}$ in a way that $\overline{MF}(x)$ be continuous and $0 < \sup_S \overline{MF}(x) < \frac{\epsilon}{2}$. It is possible by a linear approximation of $\overline{MF}(x)$ to $MF(x)$ in every element of S; see Fig. 2. There exists $f(x) \in \mathcal{L}$, such that for every $z \in \mathcal{X}$, $|\overline{MF}(z) - f(z)| < \frac{\epsilon}{2}$.

On the other hand, for every $z \in \mathcal{X}$, $|MF(z) - \overline{MF}(z)| < \frac{\epsilon}{2}$, therefore, for every $z \in \mathcal{X}$, $|MF(z) - f(z)| \leq |MF(z) - \overline{MF}(z)| + |\overline{MF}(z) - f(z)| < \epsilon$. Hence, we can conclude that the theorem holds true even if $MF(x)$ equals to 1 and 0. □

See that the precedent proof is a variation of the so-called Stone theorem [17]. Unlike [14, 12, 13], we do not have to approximate using subintervals, neither do we use other operators besides max, min or negation $N(x) = 1 - x$, to maintain the semantics of the results. □

Remark 3 For application purposes, it is enough to consider that $\mathcal{X}$ be compact (closed and bounded). It is very unusual to model real-life variables x, such that x is near to $-\infty$ or $+\infty$. Besides, compact intervals could be defined by two finite extrema as near as possible to $-\infty$ or $+\infty$.

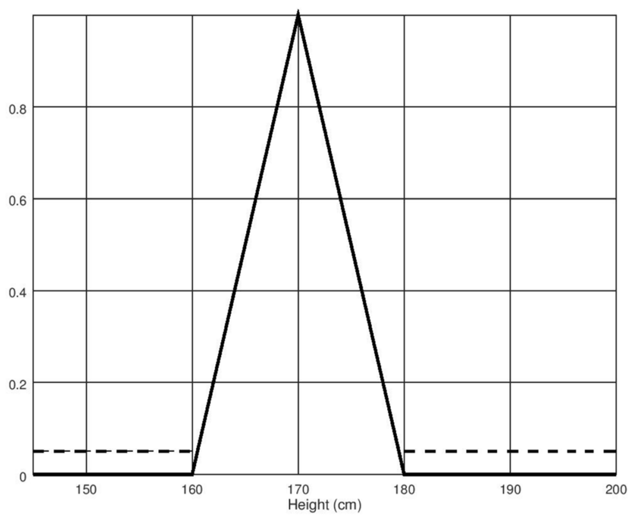


Fig. 2 Triangular sigmoidal function (solid line) and piece-wise approximation (dashed line)

Remark 4 For the sake of clarity, we will substitute $\mathcal{G}(\mathcal{X};\alpha,\gamma,\{1,\frac{1}{2}\})$ by $\mathcal{G}(\mathcal{X};\alpha,\gamma,m)$ $= \{\mathrm{NGCLV}_T(x;\alpha,\gamma,m,1)|x\in\mathcal{X},\alpha\in[a_\alpha,b_\alpha],$ $\gamma\in[a_\gamma,b_\gamma],m\in[0,1]\}$ in the theorem. The former is a subset of the latter, and it is easy to see that the theorem does not change its conclusions.

Remark 5 This theorem states the potential applicability of the NGCLVs to approximate continuous membership functions just using logical operators, like the strong negation $N(x)=1-x$ (NOT), the biggest t-norm min (AND), and the smallest t-conorm max (OR). Note that they can approximate NGCLVs based on other t-norms. This could yield a semantic approach to Fuzzy Inference Systems (FIS) [18] or to interpretable Neural Networks [19, 20]. These t-norms and t-conorms are associated with compensatory operators; see [21], which enrich the applicability of this approach.

Remark 6 Theorem 1 is valid for discontinuous $MF(x)$, with a finite number of jump discontinuities existing over a compact set. Every jump discontinuity can be approximated linearly and then the theorem is applied.

We shall state the principle, so-called *principle of representation of linguistic variables*, which is one of the cornerstones of this study and asserts the following:

Let $\mathcal{W}$ be a linguistic variable over a continuous variable set $\mathcal{X}$. Every continuous fuzzy set in $\mathcal{W}$ can be represented by a membership function in $\mathcal{NGCLV}_T(x;\alpha,\gamma,m)$, where T is the product t-norm.

A simplified version of this principle is the following:

Given $\mathcal{W}$ a linguistic variable over a continuous variable set $\mathcal{X}$. At least the primary terms in $\mathcal{W}$ and its linguistic modifiers can be represented by membership functions in $\mathcal{NGCLV}_T(x;\alpha,\gamma,m)$, where T is the product t-norm.

According to the Zadeh's definition of linguistic variable in [22], given the name of the linguistic variable, the collection of its linguistic values and the universe of discourse, then for each linguistic value, we can determine a compatibility function which belongs to $\mathcal{NGCLV}_T(x;\alpha,\gamma,m)$, for certain $\alpha\in\mathbb{R}^*$, $\gamma\in\mathbb{R}$ and $m\in[0,1]$, T is the product t-norm.

An example of this principle can be seen in Fig. 1, in which the linguistic variable "height" is represented by three parametrized membership functions, "tall" with the sigmoidal, "medium" with the function first increasing and later decreasing and finally "short" represented by the NOT sigmoidal. Besides, other linguistic values could be defined into the set $\mathcal{NGCLV}_T(x;\alpha,\gamma,m)$, e.g., "very tall" and "very short".

Remark 7 Note that according to [10], it is enough to consider a limited number of 7 ± 2 entities to describe concepts of well-defined semantics. On the other hand, also other requirements in [10] are here fulfilled, like the normality.

This flexibility differs from the parametrized family of functions we can find in literature, [3] and [5], where every function is different from each other, but in its shapes. This is an advantage, because linguistic values can be modeled and represented by fitting four parameters, it can be by experimental data's adjustment using a method of optimization. Comparing with Drakopoulos' work [12], where a sigmoidal function is not sufficient to express a linguistic variable, we use the GCLVs and the basic fuzzy operators to approximate any continuous membership function. Hence, compound predicates exhibit semantic meanings originating from the simple ones. This is not possible when considering a piece-wise approximation.

4 The Family of General Continuous Linguistic Variables

This section aims to expose the parametric properties of the GCLVs according to Dombi's approach [2, 3].

Dombi's approach utilizes four parameters to describe the properties of the membership function, two for the interval $[a, b]$, λ denotes the *sharpness* and v is the *decision level*.

Different cases are always considered, for which the parametrized functions have basically the same shape.

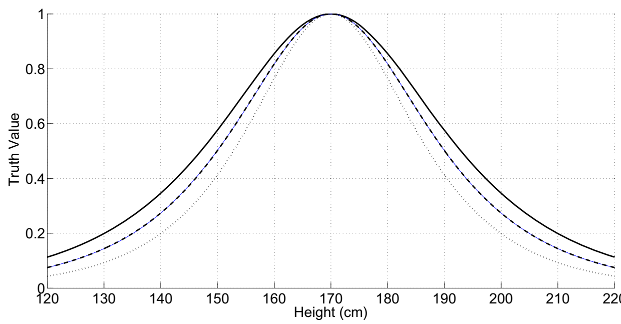


Fig. 3 Three NGCLVs representing the "Height" in cm. $\alpha = 0.1149$ was used in the function with solid line, $\alpha = 0.1313$ for the dashed line and $\alpha = 0.1532$ for the pointed line

GCLV is more general and introduces new parameters to describe other characteristics of the family.

Here, the sharpness of the general continuous linguistic variables is represented by $\alpha > 0$. This is easily justified by the differential equation $\frac{dX}{\mathrm{d}t} = \alpha X(1 - X)$, where evidently for sigmoidal membership functions, closer to 0 is α, lesser is the sharpness and vice versa. On the other hand, the NOT sigmoidal membership function is the solution for $\frac{dX}{\mathrm{d}t} = -\alpha X(1 - X)$ and satisfies the same property. Note that the sign of α represents the tendency for non-decreasing or non-increasing, while $\alpha = 0$ is a degenerated case, for which these functions are constantly equal to 0.5.

Therefore, this is also true when the GCLV is not a sigmoidal one. The degenerated case $\alpha = 0$ yields the characteristic function. See Fig. 3, where three NGCLVs representing the "height" are plotted, one from the quartet $(\beta = 130, \gamma = 170, m = 0.5, m_0 = 1)$ $(\alpha = 0.1149)$, other from $(\beta = 135, \gamma = 170, m = 0.5, m_0 = 1)$ $(\alpha = 0.1313)$ and a third one from $(\beta = 140, \gamma = 170, m = 0.5, m_0 = 1)$ $(\alpha = 0.1532)$. Let us note that the sharpest is represented there with a pointed line, which has a biggest α.

The decision level is not always easily calculated. $v = \gamma$ for the sigmoidal membership functions, i.e., these functions map γ onto 0.5.

Given a powered to m sigmoidal function,

$$v = -\frac{1}{\alpha}\ln\left(2^{\frac{1}{m}} - 1\right) + \gamma.$$

Proposition 3 *Let a SGCLV, with* $m = \frac{m_0}{2}$. *Every point* $x \in \mathbb{R}$ *has a symmetrical point with respect to* γ.

Proof Let us take $x \in \mathbb{R}$, the point $\bar{x} = \gamma + (\gamma - x)$ satisfies $\frac{1}{1+e^{-\alpha(\bar{x}-\gamma)}} = \frac{1}{1+e^{-\alpha(\gamma-x)}} = \frac{1}{1+e^{\alpha(x-\gamma)}} = 1 - \frac{1}{1+e^{-\alpha(x-\gamma)}}$ and also, $1 - \frac{1}{1+e^{-\alpha(\bar{x}-\gamma)}} = 1 - \frac{1}{1+e^{-\alpha(\gamma-x)}} = 1 - \frac{1}{1+e^{\alpha(x-\gamma)}} = \frac{1}{1+e^{-\alpha(x-\gamma)}}$. Then, to complete the proof, remain to apply the commutativity of t-norms and that the sigmoidal and the NOT sigmoidal are powered to the same exponent $\frac{m_0}{2}$. □

Particularly, let consider when NGCLV is based on the product t-norm and hence, we have to calculate the solution of the equation $\frac{X^{\frac{m_0}{2}}(1-X)^{\frac{m_0}{2}}}{M} = \frac{1}{2}$. This is equivalent to calculating the solutions of the second degree equation $X^2 - X + \left(\frac{M}{2}\right)^{\frac{2}{m_0}} = 0$ and those solutions are $X = \frac{1+\sqrt{1-4\left(\frac{M}{2}\right)^{\frac{2}{m_0}}}}{2}$ and $X = \frac{1-\sqrt{1-4\left(\frac{M}{2}\right)^{\frac{2}{m_0}}}}{2}$, therefore

$$v = \frac{1}{\alpha}\ln\left(\frac{1 + \sqrt{1 - 4\left(\frac{M}{2}\right)^{\frac{2}{m_0}}}}{1 - \sqrt{1 - 4\left(\frac{M}{2}\right)^{\frac{2}{m_0}}}}\right) + \gamma, \tag{3}$$

and

$$v = \frac{1}{\alpha}\ln\left(\frac{1 - \sqrt{1 - 4\left(\frac{M}{2}\right)^{\frac{2}{m_0}}}}{1 + \sqrt{1 - 4\left(\frac{M}{2}\right)^{\frac{2}{m_0}}}}\right) + \gamma, \tag{4}$$

respectively.

The parameter m standing in the SGCLV represents the shape of the function. We have pointed out that if $m = m_0$, it is a sigmoidal function; if $m = 0$ it is a NOT sigmoidal and if $m = \frac{m_0}{2}$, it is a symmetrical membership function. In general, $m \neq 0, m_0$ represent intermediate membership functions.

Nearer is m to 0, $0 < m < \frac{m_0}{2}$, more negative is its skewness. On the other hand, nearer is m to m_0, $\frac{m_0}{2} < m < m_0$, more positive is its skewness.

When $0 < m < \frac{m_0}{2}$, the sigmoidal function which is part of the SGCLV becomes bigger than the other part, the NOT sigmoidal and hence, the function is bigger to left than to right with respect to γ. Besides, $\frac{m_0}{2} < m < m_0$ is the opposite.

See Fig. 4, where the product t-norm was used, $\alpha = 0.1149$, $\gamma = 170$, $m_0 = 1$, from top to bottom $m = 0$, $m = 0.2$, $m = 0.5$, $m = 0.8$ and $m = 1$. Let us note the sense of the deviation with respect to γ.

SGCLVs incorporate many fuzzy concepts like hedges, which define specific aggregation operators. Considering the construction of the sigmoidal membership function where it is assumed that the marginal increase of the belief degree that "x is A" is proportional to the belief degree that "x is A" and the belief degree that "X is not A" [5], then, the SGCLV in Definition 2 results from generalizing the algebraic product by any t-norm and where the sigmoidal is modified by hedges.

For example, in Definition 2 for $C = 1$, if T is the product t-norm, $m_0 = 1$ and $m = \frac{1}{2}$, "X is A" and "X is not A" are aggregated with the geometric mean, a compensatory one. Besides, for $C = 1$, $m_0 = 2$ and $m = 1$ it is aggregated by the chosen t-norm T.

Other membership functions can be given for $m = m_0$, e.g., $m = m_0 = 2$, which means linguistically "very high". If $m = 0$ and $m_0 = 2$, it means "very low".

So far, the main advantages and properties of the GCLVs have been stated, in the following we shall explore the links of this family of membership functions with other important approaches to model the uncertainty and vagueness of the natural language.

5 Algorithms of Semantic Interpretation with Linguistic Terms

This section is devoted to expose algorithms and reflexions related to semantics expressed in form of linguistic terms, which shall be further used. These concepts are closely related to the interpretability, which constitute the initial concept to analyze in this section.

Interpretability needs further discussion because it cannot be carried out in a straightforward manner. Mencar et al. studied this subject in [23]. The main challenge of fuzzy models is that language of fuzziness should be expressed preferably in natural language or other one comprehensible to a group of human beings, namely: experts, users and designers of fuzzy models. The communication among these actors must be clear. Usually this subject is limited to fuzzy systems.

The first attempt to associate linguistic values with the aid of experts, was in [24]. This approach is valid, however not always experts are able to explain consciously how they evaluate and this task could be very difficult to achieve successfully. This limitation of expert systems becomes critical when the problem complexity increases. Therefore, other kinds of methods include a combination of expert knowledge with knowledge extracted from data, see [25], or simply, only the extraction of knowledge from data, see [26, 27]. Also, interpretable systems are frequently less accurate, which is a drawback.

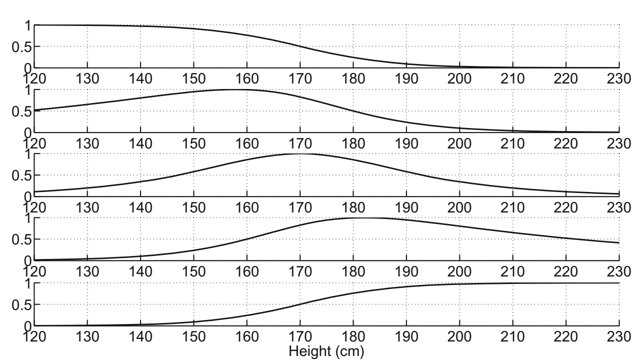


Fig. 4 Five NGCLVs representing the "Height" in cm with $\alpha = 0.1149$. From top to bottom for m = 0, m = 0.2, m = 0.5, m = 0.8 and m = 1

There exists a consensus that this problem is resolved with a fuzzy partition, see [28]. The goal is to obtain an interpretable fuzzy partition where some constraints must be satisfied, some of them coming from commonsense and others to satisfy the results of experiments in the framework of the cognitive psychology. We have made reference to the most important ones, like Distinguishability, Normality and Coverage.

The number of terms should not exceed the limit of 7 ± 2 because this is the range for the number of entities that one person could remember in its short-term memory, [11]. Nevertheless, sometimes only three is an adequate number of items, for instance, the test of Triglyceride is described as "Normal", "High" or "Very High", but never "Small" or "Very Small".

On the other hand, the Natural Zero Positioning principle will be more exact if "zero" is substituted by "neuter value".

To validate the quality of the interpretable fuzzy partition, some criteria can be seen in [29]. See also [30], in which indexes of interpretability are studied. Additionally, according to [31], human beings reduce to only one "slice" the complicated shapes of fuzzy sets, when they combine and process them. This empirical evidence has inspired the definition of distances in [32, 33].

In the following, we propose one method, which is context-dependent to the nature of problems exposed in the next section. During the exposition, we will make reference to some aspects of the previous and concise state of the art of interpretability. Before explaining the method, it is necessary to define some initial formulas and algorithms.

Let A and B be two fuzzy sets:

$$\delta_{\inf}(A,B) = \inf(A_{0.5} \cap D) - \inf(B_{0.5} \cap D), \tag{5}$$

$$\delta_{\sup}(A,B) = \sup(A_{0.5} \cap D) - \sup(B_{0.5} \cap D), \tag{6}$$

$$\eta_{\inf}(A,B) = |\delta_{\inf}(A,B)|, \tag{7}$$

$$\eta_{\sup}(A,B) = |\delta_{\sup}(A,B)|, \tag{8}$$

where $A_{0.5}$ is the 0.5-cut of A, $B_{0.5}$ is the 0.5-cut of B. The restriction of these functions to domain D prevents us from calculating with infinite 0.5-cuts, e.g., the 0.5-cuts of the sigmoidal functions are $[\gamma, +\infty[$. Thus, further in the examples it is stated $D = \left[\frac{-h}{2}, 100 + \frac{h}{2}\right]$, where $h \in]0, 100[$ and the domain is restricted to a finite interval.

We defined these functions to mimic the actual human behavior, when they deal with fuzzy sets, according to criteria in [31] that we explained above, they are simpler than those defined in [32, 33].

Given a fuzzy partition $FP^j = \{\mu_1^j, \mu_2^j, \ldots, \mu_n^j\}$ and a fuzzy set R corresponding to attribute j.

Let us define

$$\underline{i} = \left\{\text{the smallest i}, 1 \leq \text{i} \leq \text{n} | \eta_{\inf}(\text{R}, \mu_\text{i}^\text{j}) \text{ is a minimum}\right\},$$

and

$$\overline{i} = \left\{\text{the smallest i}, 1 \leq \text{i} \leq \text{n} | \eta_{\sup}(\text{R}, \mu_\text{i}^\text{j}) \text{ is a minimum}\right\}.$$

Additionally, let an index $\underline{\underline{i}}$ such that

$$\eta_{\inf}\left(R, \mu_{\underline{\underline{i}}}^j\right) = \min\left(\eta_{\inf}\left(R, \mu_{\underline{i}-1}^j\right), \eta_{\inf}\left(R, \mu_{\underline{i}+1}^j\right)\right),$$

$\underline{i} \neq 1$. When $\underline{i} = 1$ and $\delta_{\inf}\left(R, \mu_1^j\right) > 0$, $\underline{\underline{i}} = 2$, else, $\underline{\underline{i}} = 0$.

Index $\overline{\overline{i}}$ is such that

$$\eta_{\sup}\left(R, \mu_{\overline{\overline{i}}}^j\right) = \min\left(\eta_{\sup}\left(R, \mu_{\overline{i}-1}^j\right), \eta_{\sup}\left(R, \mu_{\overline{i}+1}^j\right)\right),$$

$\overline{i} \neq n$. If $\overline{i} = n$ and $\delta_{\sup}\left(R, \mu_n^j\right) < 0$, $\overline{\overline{i}} = n - 1$, else, $\overline{\overline{i}} = 0$.

Let us note that because of both, $A_{0.5} \cap D$ and $B_{0.5} \cap D$ are generally intervals then $\eta_{\inf}(A, B)$ is the distance between the lower limits of these two intervals and $\eta_{\sup}(A, B)$ is the distance between their upper limits. Thus, given R a NGCLV and FP^j a fuzzy partition of NGCLVs, we have $\underline{i}$ is the index of the element in FP^j satisfying it is the nearest one to R according to $\eta_{\inf}(R, \mu_i)$, whereas, $\overline{i}$ is the nearest one with respect to $\eta_{\sup}(R, \mu_i)$. On the other hand, $\underline{\underline{i}}$ determines the second nearer element to R in FP^j . Equivalently, $\overline{\overline{i}}$ calculates the index corresponding to $\overline{i}$. To avoid any indefiniteness when indexes $\underline{i}$ or $\overline{i}$ are extreme values like 1 or n, then we directly assign values to $\overline{\overline{i}}$ and $\underline{\underline{i}}$ including 0 to meaning the index is outside of the scope of the fuzzy partition.

Here we define two measures to be used in further investigations, they are:

$$\Delta_{\inf} = \begin{cases} \dfrac{\delta_{\inf}\left(R, \mu_{\underline{i}}^j\right)}{\eta_{\inf}\left(\mu_{\underline{i}}^j, \mu_{\underline{\underline{i}}}^j\right)} & \text{if } \underline{\underline{i}} \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

and

$$\Delta_{\sup} = \begin{cases} \dfrac{\delta_{\sup}\left(R, \mu_{\overline{i}}^j\right)}{\eta_{\sup}\left(\mu_{\overline{i}}^j, \mu_{\overline{\overline{i}}}^j\right)} & \text{if } \overline{\overline{i}} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

$\Delta_{\inf}, \Delta_{\sup} \in [-0.5, 0.5[$. We were inspired by the concept of symbolic translation introduced in the well-known 2-tuple method; see [34].

$$D_n = \frac{1}{n(n-1)} \sum_{k,l=1,2,\ldots,n;k \neq l} d(\mu_k^j, \mu_l^j). \tag{11}$$

This is the sum of dissimilarities or distances between every pair of fuzzy sets in FP^j, where

$$d(\mu_k^j, \mu_l^j) = \frac{1}{N} \sum_{q=1,2,\ldots,N} \left|\mu_k^j(x_q^j) - \mu_l^j(x_q^j)\right|. \tag{12}$$

N is the number of elements in the database. This is a measure of dissimilarity as the opposite of the measure of similarity based on the Łukasiewicz bi-implication according to the approach in [35]. These distances are basically the same of those used in [27].

Let $\mu_k^j(x) = \text{NGCLV}(x; \alpha_k^j, \gamma_k^j, m_k^j)$ and $\mu_{k+1}^j(x) = \text{NGCLV}(x; \alpha_{k+1}^j, \gamma_{k+1}^j, m_{k+1}^j)$ be two consecutive fuzzy sets in the current set of terms of the fuzzy partition. The Algorithm of merging consists in the following.

Algorithm of merging

1. Calculate the weight over the database of both fuzzy sets $w_m^j = \sum_{q=1,2,\cdots,N} \mu_m^j(x_q^j)$, $m = k, k+1$.
2. Let $y_k^j = core(\mu_k^j)$ and $y_{k+1}^j = core(\mu_{k+1}^j)$. The core of the membership functions are the x_{max} calculated with the formula shown in Prop. 2.
   Calculate $\widehat{y}_k^j = \frac{w_k^j y_k^j + w_{k+1}^j y_{k+1}^j}{w_k^j + w_{k+1}^j}$.
3. Calculate $z_1^j = \inf\left\{x \in D | \mu_k^j(x) \geq 0.5\right\}$ and $z_2^j = \sup\left\{x \in D | \mu_{k+1}^j(x) \geq 0.5\right\}$.
4. The result of merging is the interpolation of $NGCLV(x; \overline{\alpha}, \overline{\gamma}, \overline{m})$ over the three points in the plane $(\widehat{y}_k^j, 1)$, $(z_1^j, 0.5)$ and $(z_2^j, 0.5)$.

Let us note that when we merge the fuzzy sets, Distinguishability, Normality, Coverage and $hgt(\mu_i \cap \mu_{i \pm 1}) = \frac{1}{2}$ are still fulfilled. Let us observe that this equation means that the height of the intersection of two successive fuzzy sets is $\frac{1}{2}$; see [26]. One remarkable method can be found in [36]. To merge fuzzy sets is an usual practice to obtain interpretable fuzzy partitions.

See that steps 1 and 2 are defined in [27] and we adapted to NGCLV the other steps, where the former used triangular membership functions. For step 2, let us recall that y_k^j and y_{k+1}^j exist and the equations of y_m^j, $m = k, k+1$ correspond to $x_{\max}$ in Proposition 2.

For step 3, let us note that the limits of both 0.5-cuts can be calculated with the formula $x_{1,2} = \gamma \pm \frac{1}{\alpha} arccosh(7)$ only if $m = 0.5$. When $m \neq 0.5$, these calculi must be carried out numerically. We recommend for this case, to estimate the two fixed points, X_1 from equation $X_1 = \left(\frac{M}{2}(1-X_1)^{m-1}\right)^{1/m}$ and X_2 from $X_2 = 1 - \left(\frac{M}{2}X_2^{-m}\right)^{\frac{1}{1-m}}$, where M is such expressed in Proposition 2.

The iterative process is designed as follows:

1. Fix $X^0 \in [0,1]$, e.g. $X^0 = 0$ and $Error = \epsilon$, e.g. $\epsilon = 10^{-7}$.
   Let us define two functions:
   $$g_1(X) = \left(\frac{M}{2}(1-X)^{m-1}\right)^{1/m} \tag{13}$$
   and
   $$g_2(X) = 1 - \left(\frac{M}{2}X^{-m}\right)^{\frac{1}{1-m}} \tag{14}$$
2. Calculate $X_1^{prev} = X_2^{prev} = X^0$,
   $X_1^{next} = g_1(X_1^{prev})$, and $X_2^{next} = g_2(X_2^{prev})$.
3. While$(abs\,(X_1^{prev} - X_1^{next}) > Error)$
   $X := X_1^{next}$
   $X_1^{next} := g_1\,(X_1^{next})$ and $X_1^{prev} := X$.
4. While$(abs\,(X_2^{prev} - X_2^{next}) > Error)$
   $X := X_2^{next}$
   $X_2^{next} := g_1\,(X_2^{next})$ and $X_2^{prev} := X$.
5. $X_1 = X_1^{next}$ and $X_2 = X_2^{next}$. Finish.

The values we want to estimate are $x_{1,2} = \gamma - \frac{1}{\alpha} ln\left(\frac{1-X_{1,2}}{X_{1,2}}\right)$.

On the other hand, in step 4 the interpolation is performed with the help of any optimization algorithm (genetic algorithm, hill climbing algorithm, among others), to estimate $\overline{\alpha} \in [\alpha_1, \alpha_2]$, $\overline{\gamma} \in [\gamma_1, \gamma_2]$, and $\overline{m} \in [0,1]$, such that $dist^j = \left((F_1(\overline{\alpha}, \overline{\gamma}, \overline{m}) - 1)^2 + (F_2(\overline{\alpha}, \overline{\gamma}, \overline{m}) - 0.5)^2 + (F_3(\overline{\alpha}, \overline{\gamma}, \overline{m}) - 0.5)^2\right)^{1/2}$ is a minimum.

Where $F_1(\overline{\alpha}, \overline{\gamma}, \overline{m}) = \mathrm{NGCLV}(\overline{y}_k^j; \overline{\alpha}, \overline{\gamma}, \overline{m})$, $F_2(\overline{\alpha}, \overline{\gamma}, \overline{m}) = \mathrm{NGCLV}(z_1^j; \overline{\alpha}, \overline{\gamma}, \overline{m})$, and $F_3(\overline{\alpha}, \overline{\gamma}, \overline{m}) = \mathrm{NGCLV}(z_2^j; \overline{\alpha}, \overline{\gamma}, \overline{m})$.

To assign a linguistic phrase to the fuzzy set related to one attribute, we designed the following algorithm:

Algorithm of translation a fuzzy set to a linguistic phrase

1. Design *a priori* one set of terms with their semantics for the j-th attribute, $FP^j = \{\mu_1(x), \mu_2(x), \cdots, \mu_{\widehat{n}}(x)\}$, i.e., a set of membership functions, each of them with an associated linguistic value. We recommend to fix $\widehat{n} = 5$ or $\widehat{n} = 3$.
2. Given $R^j(x)$, the fuzzy set related to attribute j. We aims to associate a linguistic phrase to $R^j(x)$.
   Let us call t_{inf}^j the linguistic value associated with the fuzzy set $\mu_{\underline{i}}$ in the set of terms, and t_{sup}^j the associated one to $\mu_{\overline{i}}$.
   Now the assigment of a linguistic phrase to R is as follows:
   (a) If $\underline{i} = \overline{i}$, associate with R the linguistic value t_{inf}^j.
   (b) If $1 < \underline{i} < \overline{i} < \widehat{n}$, associate with R the linguistic phrase 'at least t_{inf}^j and at most t_{sup}^j'.
   (c) If $\underline{i} = 1$ and $\overline{i} < \widehat{n}$, associate with R the linguistic phrase 'at most t_{sup}^j'.
   (d) If $\underline{i} > 1$ and $\overline{i} = \widehat{n}$, associate with R the linguistic phrase 'at least t_{inf}^j'.
   (e) If $\underline{i} = 1$ and $\overline{i} = \widehat{n}$, do not include the attribute j in the final linguistic expression.
   The result is not interpretable.
3. Calculate Δ_{inf}^j and Δ_{sup}^j using Eq. 9 and 10, respectively, which represent the symbolic translation of the 2-tuple method.
   Output the pair of 2-tuples $\left(t_{inf}^j, \Delta_{inf}^j\right)$ and $\left(t_{sup}^j, \Delta_{sup}^j\right)$, where t_{inf}^j means the minimum linguistic term in FP^j to represent $R^j(x)$, whereas t_{sup}^j means the maximum.

Step 1 should be considered carefully. If experts use the term "Normal" instead of "Medium" in the context of the situation and also "Small" does not make sense, it is preferable to select $\widehat{n} = 3$ and linguistic values {"Normal", "High", "VeryHigh"}, let us recall the example of the Triglycerides.

If otherwise, experts use the term "Normal" instead of "Medium" and "Small" make sense, then we recommend to fix $\widehat{n} = 5$ and the linguistic values {"VerySmall", "Small", "Normal", "High", "VeryHigh"}. An example is the linguistic variable "Height", which in a medical context of endocrine disorders, the term "Normal" makes sense.

However, in the census of the population's height, {"VerySmall", "Small", "Middle", "High", "VeryHigh"} is more adequate.

The grammar developed in Step 2 is partially based on that developed in [37].

In case the user wishes to include an attribute declared as not interpretable in the precedent algorithm, we recommend to recalculate R with bigger alphas, however the cost is the diminution of the accuracy. Also, the user should consider if that fact means this attribute is irrelevant to the semantic of the predicate.

We selected the basic scheme of the method described in [27], called *hierarchical fuzzy partitioning* to create the **Algorithm to design linguistic terms**. The new proposed method is included in the first stage of the **Algorithm of translation a fuzzy set to a linguistic phrase** described above, considering that the original data are rescaled to [0, 100], and consists in the following steps:

Algorithm to design linguistic terms

1.1. Fix a value h, $h \in]0, 100[$, which is the length of the 0.5-cut of the membership functions in the initial partition for the attribute j. We recommend to choose h as small as possible. $D = \left[\frac{-h}{2}, 100 + \frac{h}{2}\right]$.
1.2. Choose the cardinality $\widehat{n}$ of the final fuzzy partition, $\widehat{n} = 3$ or $\widehat{n} = 5$.
1.3. Define an initial partition $FP^j = \left\{\mu_1^j, \mu_2^j, \cdots, \mu_n^j\right\}$, where $n = \left\lfloor \frac{100}{h} \right\rfloor + 1$, $\mu_i^j(x) = NGCLV(x; \alpha_i(h), \gamma_i(h), 0.5)$, $\gamma_i(h) = (i-1)h$ and $\alpha_i(h) = \frac{2}{h} arccosh(7)$. Let us note that FP^j satisfies Distinguishability, Normality, Coverage and $hgt\left(\mu_i \cap \mu_{i \pm 1}\right) = \frac{1}{2}$. It does not satisfy necessarily Natural Zero Positioning or $n = 7 \pm 2$.
1.4. If $n = \widehat{n}$, finish. See that now partition satisfies $n = 7 \pm 2$ or $n = 3$ keeping the other properties. Otherwise, merge every consecutive fuzzy sets, μ_i^j and μ_{i+1}^j in FP^j and evaluate D_n^i. Determine $M_n = argmax_i\left(D_n^i\right)$ and for the smallest $k \in M_n$, merge μ_k^j and μ_{k+1}^j. The other members of the partition remain as so far.
1.5. Update FP^j after the merged process, now having cardinality $n-1$.
1.6. Go to step 4.

This algorithm must be repeated for every attribute j.

Let us point out that in the step 1.1 we define $D = \left[\frac{-h}{2}, 100 + \frac{h}{2}\right]$, where h is the length of the 0.5-cut intervals of every element in FP^j. The intervals $\left[\frac{-h}{2}, 0\right]$ and $\left[100, 100 + \frac{h}{2}\right]$ are outside of [0, 100], but to including them in D guarantees these values are covered by half of the two extreme fuzzy sets of FP^j.

Therefore, in the initial partition we have D divided into n sub-intervals, each sub-interval is subset of the 0.5-cut of only one of the elements in FP^j, which means that the coverage is satisfied, i.e., every piece of data has a linguistic representation. Additionally, let us recall that $\mu_i^j(\gamma_i(h)) = 1$ because of the properties of the NGCLVs for $m = 0.5$, thus the normality is fulfilled. However, this method ignores the "Natural Zero Positioning" requirement since it is a principle defined for fuzzy systems that by their nature must contain a membership function that evaluates the zero or "nearly zero" error, [10] , which is not an objective of this tool. Finally, every membership function represents a linguistic value with a semantic "approximately $\gamma_i(h)$", this semantic changes to the one described in the **Algorithm of translation a fuzzy set to a linguistic phrase**.

During each iteration of the **Algorithm of merging**, two consecutive membership functions are merged into only one, such that the new 0.5-cut is the union of the 0.5-cuts of the merged functions; therefore, coverage is maintained. The interpolation with the pair $(y_k^j, 1)$ conserves the normality. The new membership function has the semantic "approximately y_k^j", until the algorithm finishes, and then the semantic in **Algorithm of translation a fuzzy set to a linguistic phrase** is output.

In the original hierarchical fuzzy partitioning, the algorithm starts with a clustering of data to ease the computational cost. Here we propose other manner, but the former is not excluded. We do not consider the proposed algorithm as unique; on the contrary, we recommend to analyze what is the more adequate one according to the context.

6 Illustrative Examples

In this section, we offer an illustration of how the proposed concepts can be applied to knowledge representation and how useful can be the theory we developed.

Here, as a set of data of the first example, we use a well-known problem to characterize the red wine quality by physicochemical tests; see [38, 39]. A data mining problem is resolved and our approach is consistent with this in [40–42].

The data mining in databases refers to the process of discovering useful patterns from big volume of data [40, 43]. In our approach, as in [40, 42], we aim to extract high-level knowledge, expressed in natural language from a low-level data.

Specially, we use a classification problem to illustrate the usefulness of the proposed theory. A classification problem consists in a set of examples $(x, y) \in \mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ is the feature space and $\mathcal{Y}$ is the finite label space. The objective is to develop a classifier to predict class label.

In our example, we consider a space of eleven continuous features of the red wine of the vinho verde from Portugal, see [38] and URL http://www.ics.uci.edu/mlearn/MLRepository.html. They are physicochemical tests and we want to find the model which maps them into the set of subjective quality labels. The objective is to estimate the quality of future red wines, not by experts, but by measuring physicochemical characteristics. Moreover, with our model we are able to output the results in linguistic values, which is the usual way that people express and understand the knowledge.

Example 1 This example consists of a sample of 1,599 vinho verde red wines from Portugal. There exist 11 attributes representing physicochemical tests and the last attribute represents the quality of the wine according to experts criterion. These attributes are summarized as follows:

1. Fixed acidity (g(tartaric acid)/dm^3),
2. Volatile acidity (g(acetic acid)/dm^3),
3. Citric acid (g/dm^3),
4. Residual sugar (g/dm^3),
5. Chlorides (g(sodium chloride)/dm^3),
6. Free sulfur dioxide (mg/dm^3),
7. Total sulfur dioxide (mg/dm^3),
8. Density (g/cm^3),
9. pH,
10. Sulphates (g(potassium sulphate)/dm^3),
11. Alcohol (vol.%),
12. Quality in a scale of 0 (very bad) to 10 (excellent) measured by the median of at least 3 experts criterion.

We proceed with 'linguistic mining', i.e., we express the linguistic value 'high quality' by means of the physicochemical properties given in the form of linguistic values, and we shall proceed to adjust the data of the database.

Following the order above, the notations we will use for the attributes are: F, V, C, R, Ch, Fs, T, D, P, S, A and Q.

Each datum $x_{At,i}$ corresponding to the attribute At and index i is normalized in the form $Nx_{At,i} = \frac{x_{At,i} - \min_{At,j}(x_{At,j})}{\max_{At,j}(x_{At,j}) - \min_{At,j}(x_{At,j})} \cdot 100$. In other words, the data are rescaled to the interval [0, 100].

We resolve this problem by means of the optimization problem of parameters on the following rule:

$$\begin{aligned}&(F(\mathbf{Nx}_1) \wedge V(\mathbf{Nx}_2) \wedge C(\mathbf{Nx}_3) \wedge R(\mathbf{Nx}_4) \wedge Ch(\mathbf{Nx}_5) \wedge \\ &Fs(\mathbf{Nx}_6) \wedge T(\mathbf{Nx}_7) \wedge D(\mathbf{Nx}_8) \wedge P(\mathbf{Nx}_9) \wedge S(\mathbf{Nx}_{10}) \\ &\wedge A(\mathbf{Nx}_{11})) \leftrightarrow Q(\mathbf{Nx}_{12}),\end{aligned}$$

where we refer to the notation of the attribute followed by parenthesis we mean there exists a linguistic variable representing this attribute for this predicate.

We selected this rule because of its intuitive meaning, i.e., we want to find when 'high quality' is equivalent to these physicochemical properties. It is a more restrictive version of an IF-THEN rule. However, other rules can be used and tested, furthermore, a set of rules can be tested and the one among them having the biggest truth value can be selected; see [40, 42].

According to the principle of representation of linguistic variables, these linguistic variables can be associated with NGCLVs. Hence, this rule can be also represented as:

$$\left(\bigwedge_{i=1}^{11} \mathrm{NGCLV}(\mathbf{Nx}_i; \alpha_i, \gamma_i, m_i)\right) \leftrightarrow \mathrm{NGCLV}(\mathbf{Nx}_{12}; \alpha_{12}, \gamma_{12}, m_{12}), \tag{15}$$

where $\mathrm{NGCLV}(\mathbf{Nx}_i; \alpha_i, \gamma_i, m_i)$ for $i = 1, 2, \ldots 11$ correspond to the attributes of the physicochemical properties and $\mathrm{NGCLV}(\mathbf{Nx}_{12}; \alpha_{12}, \gamma_{12}, m_{12})$ corresponds to the quality. We used the Łukasiewicz bi-implication $x \leftrightarrow y := 1 - |x - y|$ and $\wedge$ is the t-norm min. Moreover, we use the notation NGCLV and not NGCLV_T, because we understand T is the product t-norm.

We are studying the linguistic value 'high quality', therefore we fix $\gamma_{12} = 40$ (associated with quality 5 in the original scale). Let us recall that this has the truth value of 0.5, whereas $m_{12} = 1$ represents the term 'high'.

Our task is to determine the values of αs, γs and ms standing in Eq. 15, where the objective function is the arithmetic mean of this formula evaluated in every normalized element of the data set.

The problem consists on maximizing this objective function. Finally, to complete the definition of the problem, we established the constraint of the values of parameters as the following: $\alpha_i \in [0.05, 3]$ $i = 1, 2, \ldots 12$; $\gamma_i \in [5, 95]$ $i = 1, 2, \ldots 11$ and $m_i \in [0, 1]$ $i = 1, 2, \ldots 11$.

The restriction over α is heuristically justified to obtain membership functions sufficiently fuzzified, whereas the restriction over γ assures that membership functions make sense. This example aims to find linguistic values with semantics for every one of the physicochemical attributes.

Table 1 Parameters calculated for the 12 attributes of the red wine in the training set

Attribute/parameters	α	γ	m
Fixed acidity	0.0500	27.6974	0.6136
Volatile acidity	0.0529	36.6582	0.2188
Citric acid	0.0511	10.5564	0.7571
Residual sugar	0.0892	89.1991	0.0818
Chlorides	0.0515	18.8155	0.6894
Free sulfur dioxide	0.0500	5.3044	0.8486
Total sulfur dioxide	0.1193	5.0000	0.7561
Density	0.0500	7.6488	0.8160
pH	0.2110	5.2944	0.9864
Sulphates	0.1302	9.7902	0.9467
Alcohol	0.0500	26.8961	0.9088
Quality	0.0500	40.0000	1.0000

We used the optimization package in Octave 4.2.1 to calculate the optimum of the problem, specially the function sqp, which is a sequential quadratic programming solver for nonlinear problems. The results are summarized in Table 1.

The value of the objective function obtained here is 0.8908.

To check the efficacy of the method, we estimated the value of the quality for every set of objects in the test set. If the nth object in the test set has values $\overline{\mathbf{Nx}_1}$, $\overline{\mathbf{Nx}_2}$, $\overline{\mathbf{Nx}_3}$, $\overline{\mathbf{Nx}_4}$, $\overline{\mathbf{Nx}_5}$, $\overline{\mathbf{Nx}_6}$, $\overline{\mathbf{Nx}_7}$, $\overline{\mathbf{Nx}_8}$, $\overline{\mathbf{Nx}_9}$, $\overline{\mathbf{Nx}_{10}}$ and $\overline{\mathbf{Nx}_{11}}$, we apply the following steps:

1. We evaluate these values in their corresponding membership functions estimated in the training phase.
2. We estimate the real value q in [0, 100] which maximizes the predicate of the equivalence in Eq. 15, where q is evaluated in NGCLV$(\mathbf{Nx}_{12}; \alpha_{12}, 40, 1)$.
3. The value of q is rescaled to its original scale using the formula $Q = 3 + \frac{q}{20}$, where Q is the value of the quality in the interval [0, 10].
4. The estimated value is compared with the actual value in the test set. We set an absolute error tolerance τ; see [39]. For example, if $\tau = 0.25$, the estimated value is 3.1 and the actual value is 3, because $|3.1 - 3| = 0.1 \leq \tau = 0.25$, then 3.1 is classified as quality 3.

   To validate the results, Cortez et al. [39] applied the method of fivefold cross-validation, see [39, 44]. This method consists on 5 runs, where every element of the data set is used in the test set in only one of the iterations and is used in the training set for the remaining. Therefore, every element in the data set is used once in the test set and registered.

Table 2 Linguistic mining method compared with MR, NN, SVM and TFM

Method	$\tau = 0.25$	$\tau = 0.50$	$\tau = 1.00$
MR	31.2 ± 0.2	59.1 ± 0.1	88.6 ± 0.1
NN	31.1 ± 0.7	59.1 ± 0.3	88.8 ± 0.2
SVM	43.2 ± 0.6	62.4 ± 0.4	89.0 ± 0.2
TFM	39.0 ± 1.2	55.3 ± 0.4	83.5 ± 0.3
LM	39.2 ± 5.2	55.0 ± 1.7	83.4 ± 0.9

   This experiment was repeated 20 times per fold, hence they realized 100 experiments in total. They applied a t Student's test for a confidence interval with 95% of confidence level.

   To test the efficacy of the proposed method, we used the fivefold cross-validation repeated 20 times, as well. We selected to apply the fivefold cross-validation instead of the most common tenfold cross-validation, because the former one was used in [39] and here we aim to compare our approach to the methods used by Cortez et al.

   Besides, for comparison we included the Tsukamoto Fuzzy Model [18]. The results are shown in the same way as that in [39]. We restricted the values of $\alpha \in [0.05, 0.1]$ to obtain higher precision of the classification results.

   The Tsukamoto Fuzzy Model is a Fuzzy Inference System that we applied with just one IF-THEN rule. Here, we included NGCLVs instead of particular shaped membership functions.

   In Table 2, we compare the results with other methods, the Linguistic Mining (LM), with the Multiple Regression (MR), the Neural Network (NN) and the Support Vector Machine (SVM), according to [39] and for the Tsukamoto Fuzzy Model (TFM). Let us observe that the results of the proposed method is summarized in the last row, whereas both TFM and LM are based on the NGCLV. The table reports the percent of true positives depending on τ after applying five methods of classification using the fivefold cross-validation, where the error is also provided according to the t Student's test with 95% of confidence. To the best of our knowledge, so far the more accurate solution of the red wine problem appears in [39], where Cortez et al. use the Support Vector Machine method as can be seen in Table 2.

On the other hand, 0.89060 ± 0.000744 is the expected truth value for this problem.

Let us remark that for guaranteeing the accuracy of the results, it is necessary to restrict the αs to the interval [0.05, 0.1], although the ms and γs do not need to be restricted so radically. That is to say, LM and TFM are sensible to the parameters αs. This is because for classification it is necessary that two contiguous values of the attributes have different truth values, and smaller α is in GCLV more this property is fulfilled. We corroborated this fact experimentally. Nevertheless, the user can perform a rigorous sensitivity analysis of the methods we propose here with respect to parameter α.

According to Table 2, our results are comparable to the most classical ones.

The Tsukamoto Fuzzy Model gives an almost similar results of LM, except for the expected error. Our main motivation to include the Linguistic Mining is to illustrate that the theory of GCLV can be extended to other predicates beyond the classical IF-THEN rules proper of the Fuzzy Inference Systems like TFM, in this case the predicate 15.

The most important advantage over the other methods is that the result can be expressed in natural language. Next, we use methods defined in Sect. 5 for translating the precedent results to natural language. Additionally, we justify the steps of the method. Further we applied this method to the example of the red wine.

Applying the precedent method to the example of the red wine, the contribution of experts is mostly helpless because the physicochemical attributes are objective parameters.

Now, let us apply the method to the problem of red wine. Firstly, we do not include four attributes with fuzzy sets in Table 1, because their 0.5-cuts cover or almost cover completely the domain [0, 100] percent. They are, 'Citric acid' with $[-15.683, 110.582]$ as 0.5-cut, 'Chlorides' with $[-13.207, 99.930]$, 'Free sulfur dioxide' with $[-13.172, 152.947]$ and 'Density' with $[-13.841, 134.695]$. See that for Support Vector Machine model in [39], 'Citric acid', 'Density' and 'Chlorides' are the less important attributes.

To calculate the 0.5-cuts, e.g., of the 'Citric acid' with parameters $\alpha = 0.0511$, $\gamma = 10.5564$ and $m = 0.7571$, first we have $M = m^m(1-m)^{1-m} = 0.57442$. Later we calculate numerically the fixed points by iterating the equations $X_1 = \left(\frac{M}{2}(1-X_1)^{m-1}\right)^{1/m}$ and $X_2 = 1 - \left(\frac{M}{2}X_2^{-m}\right)^{\frac{1}{1-m}}$ with variables X_1 and X_2, respectively; see Eqs. 13 and 14. Finally, we obtain the limits of the 0.5-cut interval by the equations $x_{1,2} = \gamma - \frac{1}{\alpha} ln\left(\frac{1-X_{1,2}}{X_{1,2}}\right)$.

A summary of the method applied to the problem is given below, where the **Algorithm to design linguistic terms** is used:

1 Design *a priori* one set of terms. Apply the **Algorithm to design linguistic terms**.
   1.1 We chose $h = 10$ for every attribute.
   1.2 We chose $\hat{n} = 5$ for every attribute. The term 'normal' seems to be not adequate in this context.
   1.3 The initial partition for every attribute is, $FP^j = \{\mu_1^j, \mu_2^j, \ldots, \mu_{11}^j\}$, $j = 1, 2, 4, 7, 9, 10, 11$, $\mu_i^j = \mathrm{NGCLV}\left(x; \frac{2}{h} arccosh(7), \gamma_i, 0.5\right)$ and $\gamma_i = (i-1)10$.
   1.4 If $n > 5$ apply the **Algorithm of merging** when necessary and other formulas and definitions. If $n = 5$ finish.
   1.5 Update FP^j, now with cardinality n-1.
   1.6 Go to step 1.4.
2. Apply the step 2 of **Algorithm of translation from fuzzy set to linguistic phrase**.
3. Calculate $\Delta_{\inf}$ and $\Delta_{\sup}$.

For the sake of simplicity, we summarize the main results. For instance, Fig. 8 describes graphically the process to merge the attribute 'Alcohol'. Each function resulting from merging is represented in bold lines. Let us note that this process conserves Distinguishability, Normality, Coverage and $hgt(\mu_i \cap \mu_{i\pm1}) = \frac{1}{2}$.

The Algorithm of merging applied in Fig. 8 is described in what follows, with the aim of illustrating this process for j = 11:

1 We selected $h = 10$ and $\hat{n} = 5$. Thus, $D = [-5, 105]$, $\alpha_i(10) = \frac{2}{10} arccosh(7) = 0.52678$, and $n = \left\lfloor \frac{100}{h} \right\rfloor + 1 = 11$. Note the smaller h is, the bigger the accuracy of the method is.
   The initial partition consists of $FP^{11} = \{\mu_1^{11}, \mu_2^{11}, \ldots, \mu_{11}^{11}\}$, where $\mu_i^{11}(x) = \mathrm{NGCLV}(x; 0.52678, \gamma_i, 0.5)$; $\gamma_1 = 0$, $\gamma_2 = 10$, $\ldots$, $\gamma_{11} = 100$.
2. $n > 5$, so the **Algorithm of merging** is applied. Then, we calculate D_n^i of Eqs. 11 and 12 corresponding to the distance between all the elements of the fuzzy partition, after one pair of consecutive members of FP^{11} is merged. Finally, the pair that maximizes D_n^i is selected, and it is the pair μ_9^{11} and μ_{10}^{11} with $D_n = 0.0048471$.
   For merging every pair, specifically μ_9^{11} and μ_{10}^{11}, we calculated the core of μ_9^{11} and μ_{10}^{11} with the formula $x_{\max} = \frac{1}{\alpha}\ln\left(\frac{m}{1-m}\right) + \gamma$ in Prop 2. Then, $y_9^{11} = \mathrm{core}(\mu_9^{11}) = 80$ and $y_{10}^{11} = \mathrm{core}(\mu_{10}^{11}) = 90$. Next, both $w_9^{11} = \sum_{q=1,2,\ldots,N} \mu_9^{11}(x_q^{11})$ and $w_{10}^{11} = \sum_{q=1,2,\ldots,N} \mu_{10}^{11}(x_q^{11})$ are calculated, where x_q^{11} are the data corresponding to this variable in the database. Then, we calculated $\hat{y}_9^{11} = \frac{w_9^{11} y_9^{11} + w_{10}^{11} y_{10}^{11}}{w_9^{11} + w_{10}^{11}} = 82.493$.

Later, we calculated the recurrent equations $X_1 = g_1(X_1)$ and $X_2 = g_2(X_2)$ by using Eqs. 13 and 14 for μ_9^{11} and μ_{10}^{11}. From μ_9^{11} we selected the lower limit of the 0.5-cut, which is $z_1^{11} = 75$ and from μ_{10}^{11} we selected the upper limit that is $z_2^{11} = 95$. Finally, we interpolate the three ordered pairs (82.493, 1), (75, 0.5) and (95, 0.5) through a NGCLV and the obtained parameters were $\bar{\alpha} = 0.33285$, $\bar{\gamma} = 78.81824$ and $\bar{m} = 0.77262$.

Now n = 10. The new function has index 9. The rest of the steps are represented in Fig. 8.

The result of the LM for the attribute 'Alcohol' is $R^{11}(x) = \mathrm{NGCLV}(x; 0.0500, 26.8961, 0.9088)$, as shown in Table 1 and thus its 0.5-cut is the interval $I_A = [14.334, 245.855]$. Additionally, the 0.5-cuts of the membership functions in the interpretable partition are: $I_1 = [-5, 15]$, $I_2 = [15, 25]$, $I_3 = [25, 45]$, $I_4 = [45, 55]$, and $I_5 = [55, 104.9995]$. Comparing the lower limit of I_A with those of I_i, we have $-5 < 14.334 < 15$ and for the upper limits $104.9995 < 245.855$. Therefore, according to Eqs. 5 and 7, $\delta_{\inf}(R^{11}, \mu_1^{11}) = 14.334 - (-5) = 19.334$, $\delta_{\inf}(R^{11}, \mu_2^{11}) = 14.334 - 15 = -0.666$, $\eta_{\inf}(R^{11}, \mu_1^{11}) = 19.334$, and $\eta_{\inf}(R^{11}, \mu_2^{11}) = 0.666$. Comparing the lower limits, I_A is nearer to I_2 than I_1 and $\underline{i} = 2$. Between $I_{\underline{i}-1} = I_1$ and $I_{\underline{i}+1} = I_3$, I_A is nearer to I_1, thus $\underline{\underline{i}} = 1$, therefore we calculated $\eta_{\inf}(\mu_1^{11}, \mu_2^{11}) = abs(-5 - 15) = 20$. Then, since Eq. 9 $\Delta_{\inf}^{11} = -\frac{0.666}{20} = -0.0333$. Similarly, we calculated $\bar{i} = 5$, $\bar{\bar{i}} = 0$ because of $\delta_{\sup}(R^{11}, \mu_2^{11}) = 245.855 - 104.9995 = 140.86 > 0$ and then $\Delta_{\sup}^{11} = 0$.

In Fig. 9, the results of Linguistic Mining are depicted in solid lines over the linguistic system, they can be observed in Figs. 5, 6 and 7 calculated from Table 1. Finally, in Table 3 it is associated every attribute with $\Delta_{\inf}$ and $\Delta_{\sup}$,

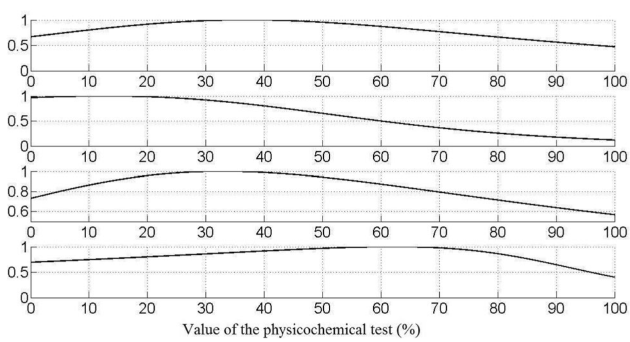


Fig. 5 Membership functions resulting from the solution of the problem, corresponding, from top to bottom, to Fixed acidity, Volatile acidity, Citric acid and Residual sugar

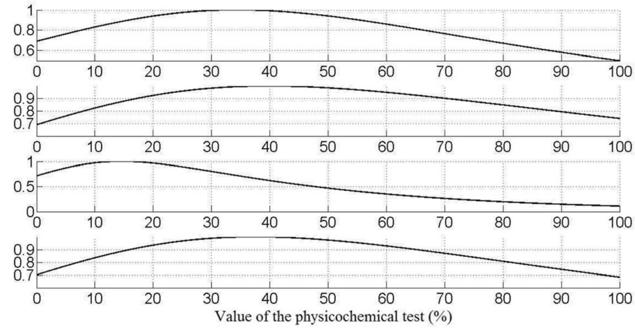


Fig. 6 Membership functions resulting from the solution of the problem, corresponding, from top to bottom, to Chlorides, Free sulfur dioxide, Total sulfur dioxide and Density

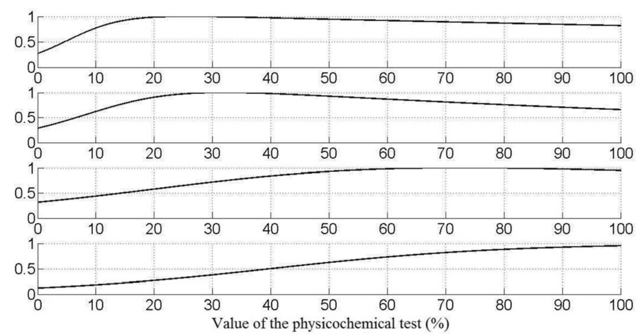


Fig. 7 Membership functions resulting from the solution of the problem, corresponding, from top to bottom, to pH, Sulphates, Alcohol and high Quality

Table 3 Labels per attribute. 'VS' is 'Very Small', 'S' is 'Small', 'H' is 'High', and 'VH' is 'Very High'

Attribute	$(\mathrm{Label}_{inf}, \Delta_{\inf})$	$(\mathrm{Label}_{sup}, \Delta_{\sup})$
Fixed acidity	(VS, 0)	(VH, − 0.17065)
Volatile acidity	(VS, 0)	(H, − 0.16827)
Residual sugar	(VS, 0)	(H, 0.068358)
Total sulfur dioxide	(VS, 0)	(S, 0.27041)
pH	(VS, 0.32060)	(VH, 0)
Sulphates	(S, 0.15837)	(VH, 0)
Alcohol	(S, − 0.033321)	(VH, 0)

whereas in Table 4 linguistic phrases are associated with attributes.

Let us remark that similarly to the 2-tuple method, the obtained values in Table 3 consist in 2-tuples with one linguistic value and one numeric symbolic translation value, which are $\Delta_{\inf}$ and $\Delta_{\sup}$. The absolute value of the symbolic translations means the displacement with respect to the fuzzy set is represented by the linguistic value, whereas the sign means the direction, namely, to left, to

Table 4 Linguistic interpretations per attribute

Attribute	Interpretation
Fixed acidity	–
Volatile acidity	'At most high'
Residual sugar	'At most high'
Total sulfur dioxide	'At most small'
pH	–
Sulphates	'At least small'
Alcohol	'At least small'

right or non-displacement. Particularly, two 2-tuples are introduced instead of only one, to represent the range of possible interpretations of the variables: e.g., the obtained result for the Alcohol is interpreted at least as 'Small' with a displacement on the left equals to 0.033321; additionally, it is at most 'Very High' with non-displacement.

Next the step 2 of the Algorithm of translation a fuzzy set to a linguistic phrase is applied. Then, revisiting the example of the Alcohol we have that $\underline{i} = 2 > 1$, which is the corresponding index for 'Small' and $\bar{i} = \hat{n} = 5$, which is the corresponding index for 'Very High', thus according to the algorithm the output is 'at least small'.

As a consequence, we have the following linguistic rule:

If 'Volatile acidity' is 'at most high' and 'Residual sugar' is 'at most high' and 'Total sulfur dioxide' is 'at most small' and 'Sulphates' are 'at least small' and 'Alcohol' is 'at least small', then 'Quality' of wine is 'High'.

Let us note that 'Fixed acidity' and 'pH' correspond to the case (e) of the Algorithm of translation a fuzzy set to a linguistic phrase. They ranged from 'VS' to 'VH', which does not express any useful information about their conditions, thus we used the symbol '–' in Table 4 to mean that the results are not interpretable.

It can be seen that because we used 0.5-cuts for calculation, the results are conservative. If we use the core of fuzzy sets instead of 0.5-cuts in precedent methods, we would obtain the better possible results per attribute, we only have to adapt the proposed method to the core of fuzzy sets.

Therefore, Eqs. 5 and 6 become in $\delta_{\inf}(A,B) = \delta_{\sup}(A,B) = A_1 - B_1$, where A_1 and B_1 are the unique values of the cores of A and B, respectively. They are calculated with the formula of $x_{\max}$ in Proposition 2. Then, Eqs. 9 and 10 convert into one single value. In the

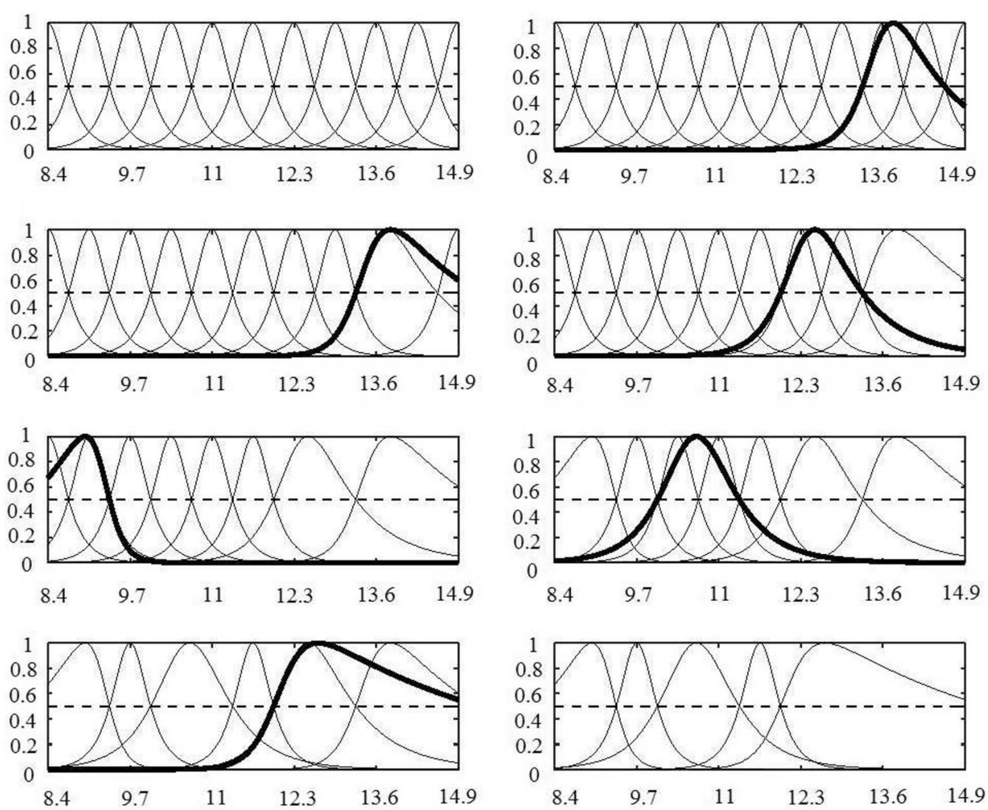


Fig. 8 Depiction of the interpretable fuzzy partition for the alcohol (on right and bottom) from an initial partition (on left and top). Functions in bold lines are calculated from merging

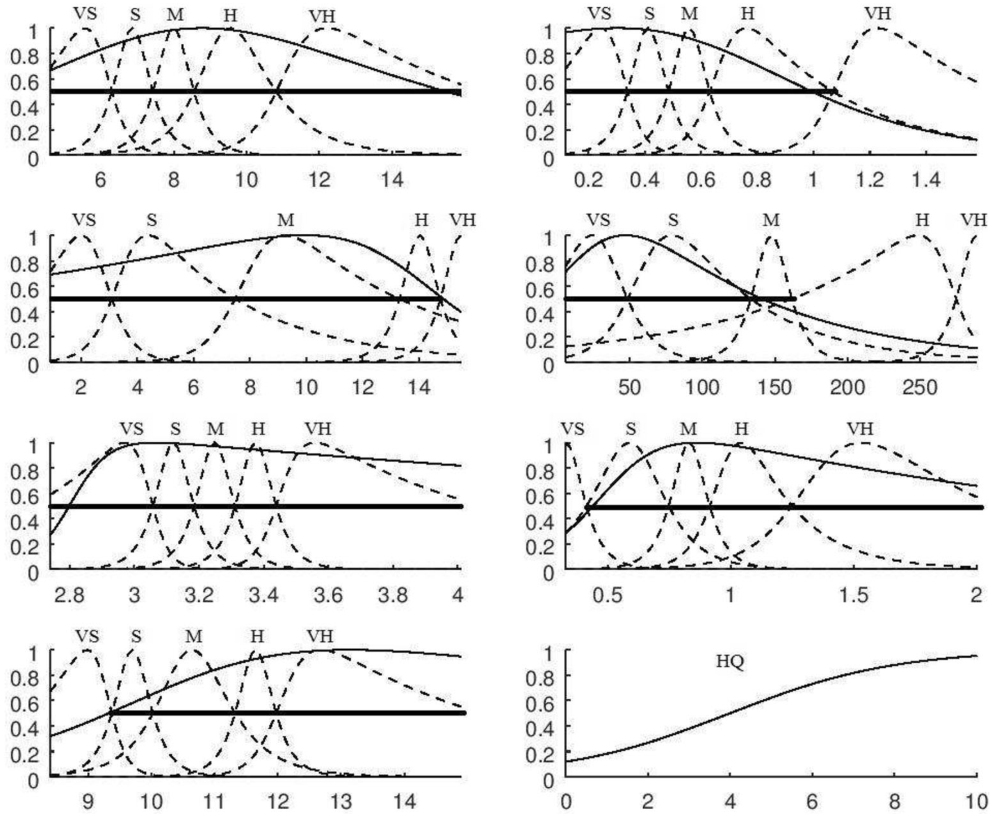


Fig. 9 Labeled interpretable fuzzy partitions (dashed lines) of, from top to bottom and from left to right, 'Fixed acidity', 'Volatile acidity', 'Residual sugar', 'Total sulfur dioxide', 'pH', 'Sulphates', 'Alcohol' and 'High quality'. LM results are in solid lines. The thick lines represent the range of functions of the fuzzy partition covered by the LM result

step 2 of the **Algorithm of translation a fuzzy set to a linguistic phrase**, $t^j_{\text{inf}} = t^j_{\text{sup}}$ and the obtained linguistic value is also unique. When $A_{0.5}$ is substituted by A_1 and $B_{0.5}$ by B_1, we lose accuracy and improve on interpretability because R^j is associated with only one element of the interpretable system.

Therefore, as a consequence of Table 5, we can say that 'Quality' of wine is the 'Highest' when 'Volatile acidity' is 'very small' and 'Residual sugar' is 'medium' and 'Total sulfur dioxide' is 'very small' and 'Sulphates' are 'medium' and 'Alcohol' is 'very high'. Let us note that now Δ is redefined for the core of the fuzzy sets and it is unique, because fuzzy sets are unimodal.

These results confirm the oenological theory, according to [39] an increase in the alcohol, a decrease of volatile acidity or a more or less high level of sulphates improve the quality of the wine, like in our conclusions. However, our approach is more informative, for example, the ideal values of alcohol, volatile acidity and sulphates are given approximately as we can see in the figures and as we expressed in natural language.

Let us remark that we have only illustrated the potentials of the precedent methods. They can be substituted or adapted according to the requirements of the problem and users.

Table 5 Best options to obtain the highest quality of wine per attribute

Attribute	(Label, Δ)
Volatile acidity	(VS, 0.35008)
Residual sugar	(M, 0.15524)
Total sulfur dioxide	(VS, 0.41952)
Sulphates	(M, 0.14829)
Alcohol	(VH, 0)

'VS' is 'Very Small', 'M' is 'Middle', and 'VH' is 'Very High'

Let us note that the statements expressed in natural language are more useful and expressive for generalization and understanding than the black box models such as Support Vector Machines, Multiple Regression and Neural Networks studied in [39].

Example 2 This example is dedicated to illustrating the application of NGCLVs in the prediction of Gas Furnace behavior in the time series corresponding to the variable X that measure the gas rate in cubic feet per minute, as it can be seen in Box et al. see [45].

For the solution we designed a two-rule Fuzzy Inference System, where we denote by G (t) the membership function of the gas rate in cubic feet per minute for time t. The IF-THEN rules are as follows:

Rule 1: IF G(t-2) is A_1 AND G(t-1) is A_2 AND G(t) is A_3 THEN G(t+1) is C_1,
Rule 2: IF G(t-2) is B_1 AND G(t-1) is B_2 AND G(t) is B_3 THEN G(t+1) is C_2,

where A_1, A_2, A_3, B_1, B_2, and B_3 are NGCLVs, which depend on the triples of parameters $(\alpha_{A_1}, \gamma_{A_1}, m_{A_1})$, $(\alpha_{A_2}, \gamma_{A_2}, m_{A_2})$, $(\alpha_{A_3}, \gamma_{A_3}, m_{A_3})$, $(\alpha_{B_1}, \gamma_{B_1}, m_{B_1})$, $(\alpha_{B_2}, \gamma_{B_2}, m_{B_2})$, and $(\alpha_{B_3}, \gamma_{B_3}, m_{B_3})$. C_1 depends on $(\alpha_{C_1}, 0, 1)$ and C_2 depends on $(\alpha_{C_2}, 0, 0)$, to mean "HIGH gas rate" and "LOW gas rate", respectively.

To find these NGCLVs, we applied the Tsukamoto FIS method, and the parameters of the NGCLVs were calculated using the sqp function of Octave 4.2.1 restricting $\alpha \in [0.05, 3]$, $\gamma \in [-2.716, 2.834]$, and $m \in [0, 1]$. We did not pre-process the data to preserve the 0 as the neutral value.

We selected the first 193 quartets as training set and the remaining 97 as a test set, which were compared with the real values. The mean absolute error (MAE) was used as error. The results are summarized in Table 6.

Table 6 Parameters of the trained Tsukamoto FIS method

Attribute/parameters	α	γ	m
A_1	0.050000	0.918379	0.255712
A_2	0.146526	− 0.475558	0.762909
A_3	0.209266	− 0.983681	0.815305
C_1	0.166896	0	1
B_1	0.058777	0.294116	0.619224
B_2	0.247053	− 0.036347	0.501238
B_3	0.227709	− 0.141662	0.436818
C_2	0.367199	0	0

Let us give a more detailed approach to explain the Tsukamoto FIS method applied to this problem. Let X_k, X_{k+1} and X_{k+2} be three successive measured values of the gas rate, taking the parameters α, γ and m, corresponding to A_1, A_2, A_3, C_1, B_1, B_2, B_3, and C_2, according to Table 6. We have to forecast the value of the gas rate X_{k+3}. For this end, $a_1 = A_1(X_k)$, $a_2 = A_2(X_{k+1})$, $a_3 = A_3(X_{k+2})$, $b_1 = B_1(X_k)$, $b_2 = B_2(X_{k+1})$, $b_3 = B_3(X_{k+2})$ are calculated.

The next step is to find $\overline{a} = \min\{a_1, a_2, a_3\}$ and $\overline{b} = \min\{b_1, b_2, b_3\}$. The predicted value of X_{k+3} is calculated by the formula $\widehat{X}_{k+3} = \frac{\overline{a}C_1^{-1}(\overline{a}) + \overline{b}C_2^{-1}(\overline{b})}{\overline{a}+\overline{b}}$, where $C_1^{-1}(\cdot)$ and $C_2^{-1}(\cdot)$ are the inverse functions of C_1 and C_2, respectively. We calculated the parameters such that the mean of distances between $\widehat{X}_{k+3}$ and the actual X_{k+3} for every value in the data set is a minimum. These parameters are those obtained in Table 6.

Membership functions are plotted in Fig. 10, where the membership functions corresponding to A_1, A_2, A_3 on top and B_1, B_2, and B_3 on bottom are drawn to the left. The membership functions of the conclusions C_1 and C_2 are on the right. The MAE obtained from comparing the predicted results with the real ones was 0.18268. Let us see Fig. 11, where the time series predicted values are depicted with dotted lines and the real values are depicted with solid lines.

This result was compared with a traditional statistical method that is the linear autoregression, in this case the Adaptive Autoregression; see details in [46]. For this, the aar function of Octave 4.2.1 was used of the tsa package, with model order parameters [3, 0], and Mode [1, 2].

In addition, the nnet-0.1.13 package was used to model Artificial Neural Networks in Octave 4.2.1. A hidden neuron and an output neuron were used. For training,"tansig" was used to represent the Tansig transfer function of ith layer, "trainlm" for backpropagation network training function. Table 7 shows MAEs of every method. Let us observe that the estimation here proposed is approximately equal to that of ANN.

This example serves to illustrate some advantages for using NGCLVs, which are the following:

- *Simplicity* The design consisted of only two rules, each of them contains only three premises. If a pre-trained offline Adaptive Neuro-Fuzzy Inference System (ANFIS) is applied, it is necessary to have more membership functions for each premise and for each rule, which means more parameters to estimate, and thus more computational time to invest; see [18].
- *Accuracy* Comparing with non-interpretable methods, such as ANN or AAR, errors are sufficiently comparable with the most accurate of them.
- *Interpretability* The **Algorithm of translation a fuzzy set to a linguistic phrase** guarantees that a linguistic label can be associated with the obtained results.

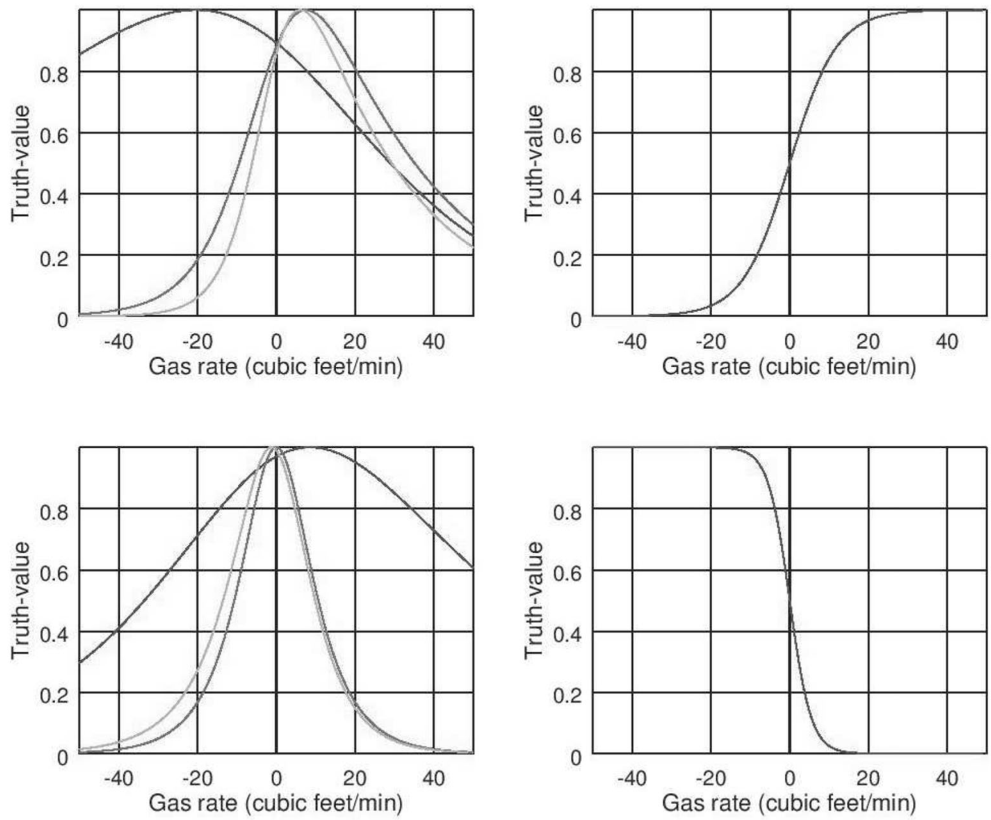


Fig. 10 Two-rule Fuzzy Inference System representing gas rate per time. The three-premise membership-functions are on the left, conclusions membership-functions are on the right. The first rule is on top, the second rule is on bottom

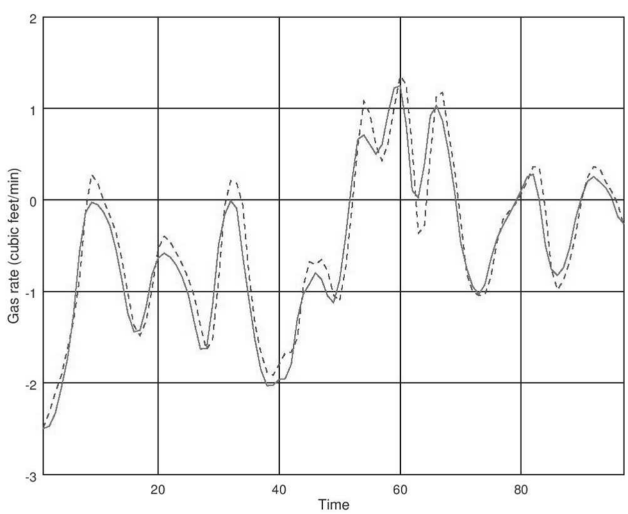


Fig. 11 Predicted and actual values of gas rate per time. Actual values are represented with solid lines, predicted values are represented with dashed lines

- *Versatility* They can be used to solve classification problems, as well as prediction. No other type of Data Mining problem is ruled out. They can be used in FIS with one or more rules.

Table 7 MAE of two traditional methods, Adaptive Autoregression and Neural Networks, and Tsukamoto Fuzzy Method based on NGCLVs

Method	MAE
AAR	0.74213
ANN	0.16510
TFM	0.18268

7 Concluding Remarks

In this paper, we have presented the study of general continuous linguistic variables. A number of main conclusions could be drawn:

We formulated the principle of representation of linguistic variables, which asserts that every family of NGCLV can be associated with a linguistic variable, because it contains differently shaped membership functions. That is to say, the main linguistic values in the linguistic variable can be represented by a family of NGCLV and to do this we have to fix m_0 and vary the values of the other three parameters. This flexibility is advantageous in the context of problems of Data Mining.

We demonstrated that every continuous membership function can be approximated by members of the family or formulas based on the operators of min, max and negation to maintain the semantics. We worked with $m_0 = 1$, $\alpha > 0$, $\gamma \in \mathbb{R}$ and $m \in [0, 1]$ and the product t-norm. The proof was based on a variant of the Stone theorem.

We illustrated the applicability of the proposed theory in Data Mining and prediction. We showed the relationship of NGCLV with the Dombi's theory. In future works we will study theoretical relationships between NGCLVs and type-2 fuzzy sets, as well as an in-depth research of potential areas of application of this theory.

Compliance with Ethical Standards

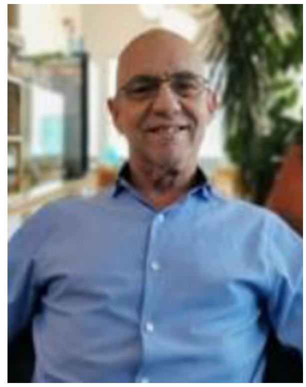
Conflict of interest The authors declare that they have no conflict of interest.

References

1. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
2. Dombi, J.: A general class of fuzzy operators, the de morgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. Fuzzy Sets Syst. **8**, 149–163 (1982). https://doi.org/10.1016/0165-0114(82)90005-7
3. Dombi, J.: Membership function as an evaluation. Fuzzy Sets Syst. **35**, 1–22 (1990)
4. Bilgiç, T., Türkşen, I.B.: Measurement of membership functions theoretical and empirical work Fundamentals of fuzzy sets. Springer, Berlin (2000)
5. Medasani, S., Kim, J., Krishnapuram, R.: An overview of membership function generation techniques for pattern recognition. Int. J. Approx. Reason. **19**, 391–417 (1998)
6. Chung, J.H., Pak, J.M., Ahn, C.K., You, S.H., Lim, M.T., Song, M.K.: Particle filtering approach to membership function adjustment in fuzzy logic systems. Neurocomputing **237**, 166–174 (2017)
7. Homaifar, A., McCormick, E.: Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. IEEE Trans. Fuzzy Syst. **3**, 129–139 (1995)
8. Wang, C.-H., Wang, W.-Y., Lee, T.-T., Tseng, P.-S.: Fuzzy b-spline membership function (bmf) and its applications in fuzzy-neural control. IEEE Trans. Syst. Man Cybern. **25**, 841–851 (1995)
9. Guo, H., Wang, X., Wang, L., Chen, D.: Delphi method for estimating membership function of uncertain set. J. Uncertain. Anal. Appl. **4**, 3 (2016)
10. Valente-de Oliveira, J.: Semantic constraints for membership function optimization. IEEE Trans. Syst. Man Cybern. **29**, 128–138 (1999)
11. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**(2), 81–97 (1956)
12. Drakopoulos, J.A.: Sigmoidal theory. Fuzzy Sets Syst. **76**, 349–363 (1995)
13. Drakopoulos, J.A., Hayes-Roth, B.: tfpr: A fuzzy and structural pattern recognition system of multi-variate time-dependent pattern classes based on sigmoidal functions. Fuzzy Sets and Systems **99**, 57–72 (1998)
14. Cybenko, G.: Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst. **2**, 303–314 (1989)
15. Narayanan, S.J., Paramasivam, I., Bhatt, R.B., Khalid, M.: A study on the approximation of clustered data to parameterized family of fuzzy membership functions for the induction of fuzzy decision trees. Cybern. Inf. Technol. **15**, 75–96 (2015)
16. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers, New York (2000)
17. Bartle, R.G.: The elements of real analysis. Wiley, New York (1964)
18. Jang, J.-S.R., Sun, C.-T., Mizutani, E.: Neuro-fuzzy and Soft Computing. A computational approach to learning and machine intelligence. Prentice Hall Inc, New York (1997)
19. Pedrycz, W., Reformat, M., Li, K.J.: Or/and neurons and the development of interpretable logic models. IEEE Trans. Neural Netw. **17**, 636–658 (2006)
20. Reyes-Galaviz, O.F., Pedrycz, W.: Fuzzy relational structures: Learning alternatives for fuzzy modeling. In IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS), pp. 374–379. Joint. IEEE, (2013)
21. Espín-Andrade, R.A., González-Caballero, E., Pedrycz, W., Fernández-González, E.R.: Archimedean-compensatory fuzzy logic systems. Int. J. Comput. Intellig. Syst. **8**(sup2), 54–62 (2015)
22. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning-I. Inf. Sci. **8**, 199–249 (1975)
23. Mencar, C., Castellano, G., Fanelli, A.M.: Some fundamental interpretability issues in fuzzy modeling. Joint EUSFLAT-LFA, pp. 100–105. Springer, Berlin (2005)
24. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man-Mach. Stud. **7**, 1–13 (1975)
25. Alonso, J.M., Magdalena, L., Guillaume, S.: Hilk: a new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. Int. J. Intellig. Syst. **23**, 761–794 (2008)
26. Espinosa, J., Vandewalle, J.: Constructing fuzzy models with linguistic integrity from numerical data-afreli algorithm. IEEE Trans. Fuzzy Syst. **8**(5), 591–600 (2000)
27. Guillaume, S., Charnomordic, B.: Generating an interpretable family of fuzzy partitions from data. IEEE Trans. Fuzzy Syst. **3**(12), 324–335 (2004)
28. Bezdek, J.C., Harris, J.D.: Fuzzy partitions and relations; an axiomatic basis for clustering. Fuzzy Sets Syst. **1**, 111–127 (1978)
29. Krzysztof, C.: Design of Interpretable Fuzzy Systems. Springer, Berlin (2017)
30. Alonso, J.M., Magdalena, L., González-Rodríguez, G.: Looking for a good fuzzy system interpretability index: an experimental approach. Int. J. Approx. Reason. **51**, 115–134 (2009)
31. Zwick, R., Carlstein, E., Budescu, D.V.: Measures of similarity among fuzzy concepts: a comparative analysis. Int. J. Approx. Reason. **12**, 221–242 (1987)
32. Koczy, L., Hirota, K.: Ordering, distance and closeness of fuzzy sets. Fuzzy Sets Syst. **59**, 281–293 (1993)
33. Subasic, P., Hirota, K.: Similarity rules and gradual rules for analogical interpolative reasoning with imprecise data. Fuzzy Sets Syst. **96**, 53–75 (1998)
34. Herrera, F., Martínez, L.: A 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans. Fuzzy Syst. **8**, 746–752 (2000)

35. González-Caballero, E., Díaz-Vázquez, S., Espín-Andrade, R.A., Montes-Rodríguez, S.: New measures of similarity based on fuzzy implications. J. Intellig. Fuzzy Syst. **33**(6), 3493–3503 (2017)
36. Herrera, F., Herrera-Viedma, E., Martínez, L.: A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. IEEE Trans. Fuzzy Syst. **16**(2), 354–370 (2008)
37. Bodenhofer, U., Bauer, P.: A formal model of interpretability of linguistic variables. Interpretability Issues in Fuzzy Modeling, pp. 524–545. Springer, Heidelberg (2003)
38. Asuncion, A., Newman, D.: Uci machine learning repository (2007)
39. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J.: Modeling wine preferences by data mining from physicochemical properties. Decis. Support Syst. **47**, 547–53 (2009)
40. Ceruto, T., Rosete, A., Espín, R.A.: Knowledge discovery by fuzzy predicates. Soft computing for business intelligence, pp. 187–196. Springer, Berlin (2014)
41. Espín-Andrade, R.A., González, E., Pedrycz, W., Fernández, E.: An interpretable logical theory: The case of compensatory fuzzy logic. International Journal of Computational Intelligence Systems **9**, 612–626 (2016)
42. Martínez, M., Espín, R.A., López, V., Rosete, A.: Discovering knowledge by fuzzy predicates in compensatory fuzzy logic using metaheuristic algorithms. Soft computing for business intelligence, pp. 161–174. Springer, Berlin (2014)
43. Herrera, F., Carmona, C.J., González, P., del Jesús, M.J.: An overview on subgroup discovery: foundations and applications. Knowl. Inf. Syst. **29**, 495–525 (2010)
44. Kiang, M.: A comparative assessment of classification methods. Decision Support Syst. **35**, 441–454 (2003)
45. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time series analysis: forecasting and control. Prentice-Hall International Inc, New Jersey (1994)
46. Schloegl, A.: The electroencephalogram and the adaptive autoregressive model: theory and applications. Shaker, Aachen (2000)

Erick González-Caballero received his PhD degree (2013) in technical science from the Technical University of Havana in Cuba, and de BS degree (1998) in Mathematics from the University of Havana in Cuba. After graduating, he works as a scientist or teacher at various universities and research centers in Cuba. In 2011 he continued his teaching and scientific activities at the Technical University of Havana. He was a Postdoctoral Visitor (2016) at the University of Oviedo, Spain. His research interests include, decision-making under uncertainty, rough sets, fuzzy sets and their generalizations, image processing, expert systems, cooperative game theory, soft computing, among others. He has published in important journals like "Group Decision and Negotiation", "International Journal of Computational Intelligence Systems", and "Journal of Intelligent & Fuzzy Systems", also in Atlantis Press and Springer book chapters.

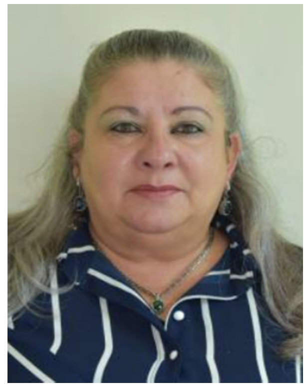


Rafael A. Espín-Andrade is Full Professor of the Institute of Multidisciplinary Research for Innovation Management and Entrepreneurial Development (IMIGIDE) of the Accounting and Administration Faculty of the Autonomous University of Coahuila, Mexico. Co-president of Eureka's Community www.eurekascommunity.net, Coordinator of the Iberian American Network of Knowledge Discovery Eureka Iberoamerica, Executive secretary of the Coordination Board of the Trans disciplinary Science Network Eureka International www.eurekainternational.org. He has been Executive secretary of the Scientific Committee of Eureka SD: Erasmus Mundus project "Enhancement for Education and Research in Areas Useful for Sustainable Development" www.eureka-sd-project. He is a Mexican National Researcher of level II (SNI II). He is graduated in Mathematics from the Havana University (1986), Master in Optimization and Decision Theory from Havana Technical University (CUJAE) 1996 and Ph.D. in Industrial Engineering from CUJAE. (2000). He made Postdoctoral Studies in Business Informatics in the Otto von Guericke University of Magdeburg, Germany in 2003. He has been co-coordinator of various research and academic projects in collaboration with the University of Oldenburg, Germany in the area of Business Intelligence, Business Informatics and Environmental Information Systems like the Bi-national Cuba – Germany Ph.D. program between the universities of Oldenburg, the Havana Technical University (CUJAE) and Central University of Las Villas (UCLV) organized by the DAAD project DEEBIS (Doctoral Education in Environmental and Business Information Systems). He leaded the consultancy and research group GEMINIS: Management in uncertainty from 1999-2005. He has been chair of several international scientific congresses, workshops and clusters of sessions. He has been professor of Decision Making Theory, Games Theory, Logics, Fuzzy Logic, Management Sciences and Business Intelligence in universities of Germany, Argentina, Bolivia, Brazil, Cuba, Mexico and Venezuela. He has published numerous papers in specialized journals and has participated as lecturer in multiples congresses, conferences and seminars in America and Europe. His research interests are Mathematical Fuzzy Logic, Knowledge Discovery, Decision Making, Cooperative Games Theory, Management Sciences, Business Analytics and Business Intelligence.

Witold Pedrycz (IEEE Fellow, 1998) is Professor and Canada Research Chair (CRC) in Computational Intelligence in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. In 2009 Dr. Pedrycz was elected a foreign member of the Polish Academy of Sciences. In 2012 he was elected a Fellow of the Royal Society of Canada. Witold Pedrycz has been a member of numerous program committees of IEEE conferences in the area of fuzzy sets and neurocomputing. In 2007 he received a prestigious Norbert Wiener award from the IEEE Systems, Man, and Cybernetics

Society. He is a recipient of the IEEE Canada Computer Engineering Medal, a Cajastur Prize for Soft Computing from the European Centre for Soft Computing, a Killam Prize, and a Fuzzy Pioneer Award from the IEEE Computational Intelligence Society.

Luis Martínez received the M.Sc. and Ph.D. degrees in computer sciences from the University of Granada, Granada, Spain, in 1993 and 1999, respectively. He is currently a Full Professor with the Computer Science Department, University of Jaén, Jaén, Spain. He is also Visiting Professor in University of Technology Sydney, University of Portsmouth (Isambard Kingdom Brunel Fellowship Scheme), and in the Wuhan University of Technology (Chutian Scholar), Guest Professor in the South west Jiaotong University and Honourable professor in Xihua University both in Chengdu (China). He has coedited eleven journal special issues on fuzzy preference modelling, softcomputing, linguistic decision making and fuzzy sets theory and has been main researcher in 16 R&D projects, also has published more than 150 papers in journals indexed by the SCI and more than 200 contributions in Inter/national Conferences related to his areas. His current research interests include decision making, fuzzy logic-based systems, computing with words and recommender systems. He was a recipient of the IEEE Transactions on fuzzy systems Outstanding Paper Award 2008 and 2012 (best owed in 2011 and 2015, respectively). He is a Co-Editor-in-Chief of the International Journal of Computational Intelligence Systems and an Associate Editor of the journals, including the IEEE Transactions on Fuzzy Systems, Knowledge Based Systems, Information Fusion, the International Journal of Fuzzy Systems, and the Journal of Intelligent & Fuzzy Systems. He is a member of IEEE and of the European Society for Fuzzy Logic and Technology. Eventually, he has been appointed as Highly Cited Researcher 2017-2019 in Computer sciences.



Liliana A. Guerrero-Ramos is Cuban by birth and Mexican by naturalization. She has a Bachelor of Library & Information Sciences, a Master of Information Management, and a Ph.D. in Information Sciences. She has 39 years of teaching experience and is a researcher in the Academic Group of Business Management of the Autonomous University of Coahuila in Mexico (Universidad Autónoma de Coahuila). One of her obligations in the University is being on charge of the Research Line called "Intelligent Management of Organizations", which fundamentally aims to generate and apply a business management methodology that applies Artificial Intelligence and Emotional Intelligence techniques, obtaining, as a result, an improvement in the Key Performance Index. She has participated in 30 national, regional, and international scientific conferences. She counts with 5 books, 8 book chapters, and 26 scientific articles in indexed journals published.