

Hierarchical Architectures of Fuzzy Models: From Type-1 fuzzy sets to Information Granules of Higher Type

Witold Pedrycz

*Department of Electrical & Computer Engineering, University of Alberta, Edmonton, T6R 2V4 Canada
and Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Received: 05-11-2009

Accepted: 28-05-2010

Abstract

Complex phenomena are perceived from different perspectives, diversified conceptual points of view and at various levels of granularity. Symbolic and sub-symbolic processing becomes an inherently visible computing practice. Distributed nature of perception becomes reflected in topologies of multi-agent systems. All of these facets challenge the well-established paradigms of system modeling including fuzzy models and neural networks. In spite of the diversity of existing architectures and underlying algorithms, a vast majority of fuzzy models adheres to the surprisingly homogeneous principles of Granular Computing, that are associated with the processing of granular information. In this study, being cognizant of this underpinning, we concentrate on the architectures and fundamentals supporting the reconciliation and characterization of a family of fuzzy models aimed at the representation of the same system (phenomenon) from different cognitive perspectives. The variety of points of view is reflected in different levels of granularity (specificity) of fuzzy sets present in individual models as well as different feature (attribute) spaces being used in the individual models. We discuss a way in which type-2 fuzzy sets come to the play as a result of the overall characterization. An effective way of determining of such fuzzy sets is presented. Further studies on the interpretability of fuzzy sets at the level of linguistic valuation are presented and with this regard where it is shown how these can be carried out in the setting of type-2 fuzzy sets. The question of logic operators constructed in presence of a large number of fuzzy sets is raised along with a proposal of statistically grounded logic operators, which capture some characteristics of membership degrees to be processed.

Keywords: fuzzy model, granular processing, type-2 fuzzy set, linguistic membership, interpretability, statistically grounded logic operations.

1. Introduction

Complex systems and phenomena are perceived from a variety of viewpoints. These perspectives depend on the domain knowledge of the observer, availability of data, interest in detail, time framework, just to name a few. In more technical terms, there are two important aspects supporting the diversity to be investigated, namely (a) characterization and modification of feature (attribute) spaces and (b) adjustment of granularity of information used in a particular model of the underlying phenomenon. Attribute space is directly implied by a suite of variables (attributes) being available to us and afterwards used in the construction of the corresponding model.

An interesting problem arises when it comes to the formation of a global view (model) when having a collection of fuzzy models associated with individual

viewpoints and constructed on a basis of the individual data, their granularity and the corresponding feature space. As a commonly encountered design practices utilized in fuzzy modeling are concerned with the formation of information granules regarded as a blueprint (skeleton) of the model, in our investigation on the reconciliation and building a unified and comprehensive view at the system we will concentrate on the manipulation of information granules. Based on the experimental evidence of information granules articulated through individual models, we discuss an optimization problem of forming a combined feature space of the highest descriptive (representation) power (which is quantified through some performance indexes).

We also concentrate on the associated issues of enhanced interpretability of fuzzy sets by elaborating on the role of type-2 fuzzy sets (which offers an effective vehicle of linguistic quantification of numeric

membership degrees) and shadowed sets (with their ability to express uncertainty). Along the line of further enhancements of logic operators used in fuzzy modeling (which tend to reveal some weaknesses in case of dealing with large number of arguments), we discuss the use of statistically grounded logic operators which take into consideration some underlying knowledge of statistically character.

The overall structure of the paper is reflective of main objectives outlined above. We start with some notes on the unified principle of fuzzy modeling, which could be easily highlighted in spite of the inherent diversity of algorithmic developments (Section 2). In Section 3, we move on to the hierarchy of perspectives and their reconciliation. The detailed algorithmic aspects are presented in Sections 4 and 5. Type-2 fuzzy sets and shadowed sets along with their interpretability aspects are discussed in Section 6. Section 7 is devoted to the class of statistically grounded logic operators.

2. The Unified Principle of Fuzzy Modeling and the Diversity of Fuzzy Models

Fuzzy models and fuzzy modeling create a surprisingly diversified algorithmic and development landscape. Since the very inception of the concept of fuzzy models (which could be traced back to the early papers of Zadeh coming in the 70s, linguistic models of Wenstop¹, fuzzy models of van Kickert², relational models of Tong³ and Pedrycz⁴, fuzzy identification and control⁵ and fuzzy processes of Kandel⁶, just to recall a handful of early developments in the realm of fuzzy modeling), we witness how the fuzzy models become more advanced and how more design activities become engaged. The highly visible trend witnessed today is concerned with the formation of neurofuzzy models (in which we rely heavily on the learning abilities provided by neurocomputing)⁷. The user friendliness aspect is also highly advocated⁸. The contributions to of Computational Intelligence and the usage of this technology to fuzzy modeling becomes more profound and is commonly encountered as we vividly start realizing the advantages of structural optimization – an important faculty supplied by evolutionary optimization (through genetic algorithms, evolutionary strategies, ant colonies, particle swarm optimization and other population-based optimization approaches). The optimization of the structure of the fuzzy models opens

a new, highly promising design avenue which has to be explored to the fullest extent.

While the diversity of architectures of fuzzy models and numerous ways in which they are constructed could be overwhelming and even somewhat confusing to some extent, there is a clearly identified denominator where are common to all these models: in one way or other fuzzy models are formed on a basis of information granules – fuzzy sets^{9,10,11,12}. Those information granules constitute a backbone, skeleton or a blueprint of the model. Once they have been formed (usually through searching for a structure in numeric data and the use of fuzzy clustering), the other more detailed components of the models are determined (viz. the values of their parameters are estimated). The granularity of fuzzy models is a fundamental feature, the architecture of specific local models, their aggregation and refinements are of secondary nature.

In what follows, we recall the two fundamental categories of fuzzy models, which serve as an excellent illustration of the generalized observation made above. While we can appreciate the evident variety of the models, by contrast the underlying commonality becomes even more pronounced.

Fuzzy rule-based models These models assume a well-known structure of rules of the form “if condition \mathbf{x} is A_i then output is $\Phi_i(\mathbf{x}, \mathbf{a}_i)$, $i = 1, 2, \dots, c$. Here A_i is an information granule while $\Phi_i(\mathbf{x}, \mathbf{a}_i)$ is a conclusion of the i -th rule with \mathbf{a}_i being a vector of adjustable (optimizable) parameters. From the modeling perspective the expression $\Phi_i(\mathbf{x}, \mathbf{a}_i)$ could be literally *anything*, say
fuzzy set,
linear or nonlinear regression function,
difference or differential equation,
finite state machine,
neural network

...

An interesting alternative could emerge if we allow for a heterogeneous nature of the local models, which naturally calls for the structural optimization and the use of evolutionary techniques.

Fuzzy associations Associations are formed for information granules and capture how they interact

between themselves. Activation of the granules – A_i s in successive discrete time moments can be governed by the expression

$$A_i(k+1) = \Pi(A_1(k), A_2(k), \dots, A_c(k), \mathbf{a}_i) \quad (1)$$

$i = 1, 2, \dots, c$. Here Π stands for the mapping between information granules and \mathbf{a}_i is a vector of the parameters of this mapping. More specifically, $A_i(k)$ should be viewed as a shorthand used to describe a level of activation of A_i caused by some input \mathbf{x} so being more formal, the expression should be read as $A_i[\mathbf{x}](k+1)$ which means that the input present at the k -th time moment gives rise to the membership degree of the i -th fuzzy set. The form of (1) could be quite diversified; in the simplest form (which is encountered in cognitive maps or fuzzy cognitive maps), we rewrite P as follows

$$A_i[\mathbf{x}](k+1) = f\left(\sum_{j=1}^c w_{ij} A_j[\mathbf{x}](k) + \tau_i\right) \quad (2)$$

The structure of the model of this nature is illustrated in Figure 1. We note that this model exhibits a fairly loose structure which is captured in the form of the dependencies between the activation levels of the nodes (fuzzy sets). The strength of linkages between the nodes is expressed as w_{ij} while τ_i is the bias associated with the mapping at node “ i ”. Interestingly, we can treat (2) as a recurrent neuron governing the dynamics of the i -th node of the structure.

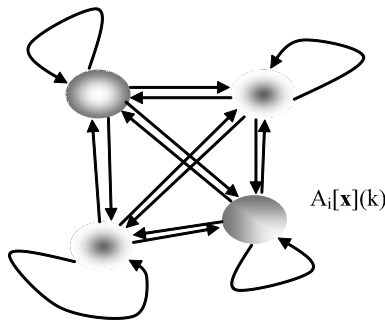


Figure 1. A web of connections between the conceptual nodes formed by the information granules and their activation levels

Information granules are the result of clustering numeric data forming the experimental evidence about the system or process. Fuzzy clustering can be realized in many different ways. Fuzzy C-Means (FCM) is in common usage. One pertinent observation is that information granules – fuzzy sets A_1, A_2, \dots, A_c are fully

described by their prototypes (note that given the family of the prototypes $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$, the membership functions of the fuzzy sets are expressed with their help in a unique manner).

$$A_i(\mathbf{x}) = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x} - \mathbf{v}_i\|}{\|\mathbf{x} - \mathbf{v}_j\|} \right)^{2/(m-1)}} \quad (3)$$

with \mathbf{x} being a certain point in the input space. Here “ m ”, $m > 1$ is an adjustable parameter using which we can modify the shape of the corresponding membership functions. The distance function is denoted by $\|\cdot\|$. Graphically, as illustrated in Figure 2, we observe that data are granulated, and detailed and local models are being formed whose relevance (usage) is restricted to the realm of the corresponding information granules. In case of fuzzy sets we are concerned with the degrees of relevance of the corresponding local models.

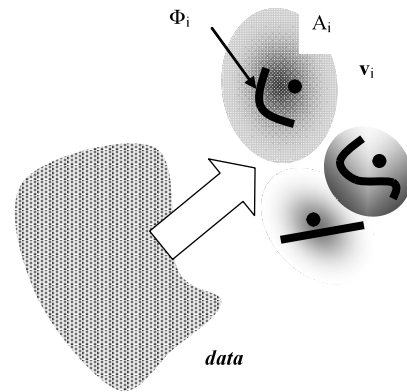


Figure 2. The framework of fuzzy modeling: data are cast in a certain perspective of information granules A_i and with associated are local models Φ_i

There is also another striking similarity between fuzzy models: all of them pertain to a single data set and a single perspective from which we perceive the phenomenon (system). Given the maturity of the field (as it has been established for several decades), we can envision a paradigm shift as we are facing with growing spectrum of ways in which data and information are gathered.

3. A Multitude of Perceptions, Development of a Hierarchy and a Construction of the Family of Granular Representatives

Let us now revisit the generic problem discussed in the previous section and place it in a far more general setting. Quite commonly the same complex phenomenon (system) can be perceived (observed) from a variety of perspectives. Different data $D[1], D[2], \dots, D[p]$ are associated with these perspectives. The diversity of the points of view (perceptions) being formed in this manner manifests in two different ways. First, there could be different feature spaces (attributes) using which we describe the system when observing it. While we admit that feature spaces F_1, F_2, \dots , etc could be different.. Second, the granularity associated with each perspective could be different and this directly quantifies a level of abstraction or a level of detail we establish when looking at the problem/data. This overall structure results in collections of information granules associated with each data. Denote the families of the corresponding information granules by S_1, S_2, \dots, S_p . The information granules (clusters) associated with each view may come with their local models (these models, however, are not considered here). Refer to Figure 3 for the details and the pertinent notation being used here.

Our objective is to determine a subset S of information granules which optimize S_1, S_2, \dots , and S_p where the optimization is sought in the sense of a certain performance index. $Q, \text{Min}_S Q$. The number of information granules forming S is equal to “ c ” and is substantially lower than the sum of information granules, that is $\text{card}(S_1) + \text{card}(S_2) + \dots + \text{card}(S_p)$.

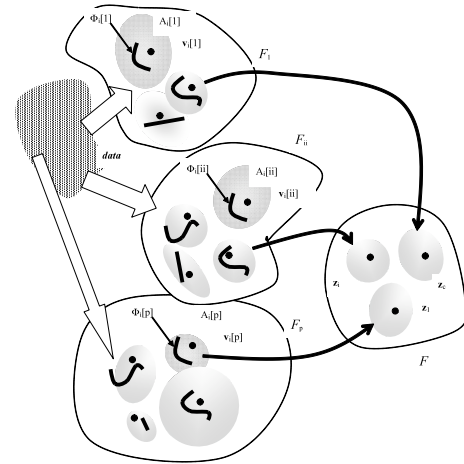


Figure 3. Hierarchical and multifaceted knowledge aggregation through selection of information granules

Let us refer to the notation present in Figure 3 using which we explain how the information granules are expressed by the representatives formed at the higher level of abstraction and how the performance index is being constructed. From now on, from the algorithmic perspective we assume that information granules are uniquely represented through their prototypes (in other words, given a collection of the prototypes, one can compute the membership functions of the corresponding fuzzy sets. In this sense the prototypes associated with a certain ii -th view $v_1[ii], v_2[ii], \dots, v_{c[ii]}[ii]$ produce the corresponding membership functions $A_1[ii], A_2[ii], \dots, A_{c[ii]}[ii]$. They are governed by the following expressions that are specific to the ii -th feature space F_{ii}

$$A_i(\mathbf{x})[ii] = \frac{1}{\sum_{j=1}^{c[ii]} \left(\frac{\|\mathbf{x} - \mathbf{v}_i[ii]\|}{\|\mathbf{x} - \mathbf{v}_j[ii]\|} \right)^{2/(m_{ii}-1)}}. \quad (4)$$

$\mathbf{x} \in F_{ii}$. The fuzzification m_{ii} coefficient is related with the feature space as well and depends on $D[ii]$. The prototypes that form the set S are denoted by $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_c$. They uniquely define fuzzy sets B_1, B_2, \dots, B_c . Here the number of information granules (c) can be adjusted on a basis of the resulting values of the performance index.

4. Representation of Fuzzy Sets by Aggregates

Any prototype $\mathbf{v}_k[ii]$ present in the set of information granules \mathbf{S}_l ($l=1,2, \dots, p$) is expressed in terms of the prototypes $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_c$. We use the well-known formula

$$u_i(\mathbf{v}_k[ii]) = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{v}_k[ii] - \mathbf{z}_i\|_{F_{ii} \cap F_{i_0}}}{\|\mathbf{v}_k[ii] - \mathbf{z}_j\|_{F_{ii} \cap F_{j_0}}} \right)^{2/(m-1)}}. \quad (5)$$

where $\mathbf{v}_k[ii] \in F_{ii}$ and $u_i(\mathbf{v}_k[ii])$ is the degree of membership (association) of $\mathbf{v}_k[ii]$ computed with respect to \mathbf{z}_i , $i=1, 2, \dots, c[ii]$, $ii=1, 2, \dots, p$; see Figure 4. In essence this representation scheme resemblance the one of encoding data for a given codebook.

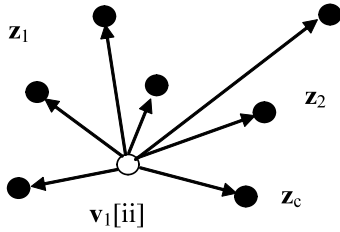


Figure 4. Representation of prototypes $\mathbf{v}_k[ii]$ from $D[ii]$ with the use of $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_c$.

Note that the distance in (5) is computed considering the intersection of the corresponding feature spaces as those spaces could be different and may vary from one collection of information granules to another. In general, we have $\mathbf{z}_i \in F_{i_0}$ and $\mathbf{z}^j \in F_{j_0}$. Recall also that the space of the subsets of information granules \mathbf{S} could be heterogeneous in the sense the elements of \mathbf{S} are reflective of different perspectives assumed when dealing with the system under discussion.

The optimized performance index Q can be formed in a way so that we emphasize the relevance of \mathbf{S} vis-à-vis $\mathbf{S}_1, \mathbf{S}_2, \dots$, and \mathbf{S}_p . Two general approaches can be distinguished:

minimum entropy (uncertainty) representation. We use the activation levels (membership grades) $u_i(\mathbf{v}_k[ii])$ and formulate the performance index in additive form by summing up the values of entropy $H(\cdot)$ computed for the individual membership grades, $H(u_i(\mathbf{v}_k[ii]))$. The summation is carried out for all \mathbf{S}_i and all elements (information granules) in \mathbf{S} ,

$$Q = \sum_{ii=1}^p \sum_{i=1}^c \sum_{k=1}^{c[ii]} H(u_i(\mathbf{v}_k[ii])). \quad (6)$$

The rationale behind the use of the entropy measure is that we would like to see such elements of \mathbf{S} so that all other elements we are intend to represent with their use are either close to them or quite remote (in both cases the associated value of entropy is low) but their distribution is such that none of elements in \mathbf{S}_i is assigned with similar membership grades (a fact which translates into higher values of entropy).

The performance index can be formed on a basis of reconstruction error, which expresses how well elements in all \mathbf{S}_i s are represented (mapped) in terms of the elements of \mathbf{S} . Formally, we express this requirement as follows

$$Q = \sum_{ii=1}^p \sum_{k=1}^{c[ii]} \|\hat{\mathbf{v}}(\mathbf{v}_k[ii]) - \mathbf{v}_k[ii]\|_{F_{ii}}^2. \quad (7)$$

where $\hat{\mathbf{v}}(\mathbf{v}_k[ii])$ is a reconstruction of $\mathbf{v}_k[ii]$ which has been completed when making use of the family $\{\mathbf{z}_i\}$. The reconstruction formula reads as follows

$$\hat{\mathbf{v}}(\mathbf{v}_k[ii]) = \sum_{i=1}^c u_i^m(\mathbf{v}_k[ii]) \mathbf{z}_i. \quad (8)$$

where $u_i(\mathbf{v}_k[ii])$ are computed in the way already introduced by means of (5), $k=1, 2, \dots, c[ii]$, $ii=1, 2, \dots, p$. The alternative expression for the reconstruction of $\mathbf{v}_k[ii]$ is expressed in the form

$$\hat{\mathbf{v}}(\mathbf{v}_k[ii]) = \sum_{i=1}^c u_i^m(\mathbf{v}_k[ii]) \mathbf{z}_i / \sum_{i=1}^c u_i^m(\mathbf{v}_k[ii]). \quad (9)$$

With regard to the reconstruction criterion, we request that it is possible to reconstruct prototypes (or data) from any feature space. This requirement states that the selected feature spaces should satisfy the “coverage” condition which means that

$$\bigcup_{k=1}^c F_{i_k} = \bigcup_{i=1}^p F_i. \quad (10)$$

From the optimization perspective, the formation of the subset of the prototypes (which in fact results in the concatenation of the existing feature spaces) is a

combinatorial problem. The use of techniques of Evolutionary Computing is a sound optimization alternative. More formally, for given value of “ c ”, we are interested in forming a collection of prototypes $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_c\}$ coming from the prototypes formed in F_1, F_2, \dots, F_p so that the performance index (6) or (7) becomes minimized. Furthermore with regard to (7) we require that the constraint (10) be satisfied. In this sense, the original optimization task is changed to the one with constraints. It reads as follows

$$\text{Min}_{\mathbf{z}} Q$$

or

$$\text{Min}_{\mathbf{z}} Q \text{ subject to (10)}$$

where Q is represented either through (6) or (7).

The fuzzification coefficient (m) is another parameter whose value could be adjusted. If we allow for this additional flexibility, we have

$$\text{Min}_{\mathbf{z}, m} Q \text{ subject to } m > 1$$

or

$$\text{Min}_{\mathbf{z}, m} Q \text{ subject to (10) and } m > 1$$

5. Information Granules as Descriptors of the Families of Numeric Prototypes

The crucial phase is to associate each of these selected (optimized) prototypes $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_c$ with the prototypes we started with at the lower processing level (viz. those coming as a result of setting up some individual perspective). Recall that the total number of prototypes in $D[1], D[2], \dots, D[p]$ is equal to $c[1] + c[2] + \dots + c[p]$. The assignment mechanism relies on the determination of the maximal association between given \mathbf{z}_i as in this way we can identify (link) the corresponding prototypes in each $D[1], D[2], \dots, D[p]$. These prototypes are those which are linked with \mathbf{z}_i to highest extent, see Figure 5.

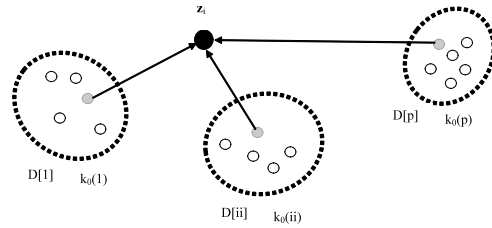


Figure 5. The determination of correspondence between \mathbf{z}_i and the prototypes in $D[1], D[2], \dots, D[p]$ identified as having the highest level of matching

More specifically, for each ii we compute the membership degrees $u_i(\mathbf{v}_k[ii])$ and find the prototype $k_0(ii)$ for which the following relationship becomes satisfied

$$k_0(ii) = \arg \max_{k=1,2,\dots,c[ii]} u_i(\mathbf{v}_k[ii]) \quad (11)$$

By completing this process for all $D[ii]$'s we end up having a collection of indexes $k_0(1), k_0(2), \dots, k_0(p)$. In terms of the individual coordinates of \mathbf{z}_i we associate with them the corresponding coordinates of the prototypes selected in this way. For the fixed coordinate, denote these values by a_0 (the corresponding coordinate of \mathbf{z}_i) and a_1, a_2, \dots, a_m , where $m \leq p$. The associated membership degrees are $\mu_1, \mu_2, \dots, \mu_m$. On this basis we can construct a granular representation of the prototype. The development of this information granule is guided by the *principle of justifiable granularity*. Given the set of pairs $(a_0, 1), (a_1, \mu_1), \dots, (a_m, \mu_m)$, Figure 6, we are interested in representing these membership values by spanning an interval $[a_1, a_2]$ around a_0 such that it realizes an intuitively appealing procedure: increase high membership values to 1 and reduce to 0 low membership values. In this sense we form an interval, see Figure 6. The formal rule of constructing the interval can be expressed in the following way

if $a_i \in [a_-, a_+]$ then elevate to membership grades to 1

if $a_i \notin [a_-, a_+]$ then reduce membership grades to 0

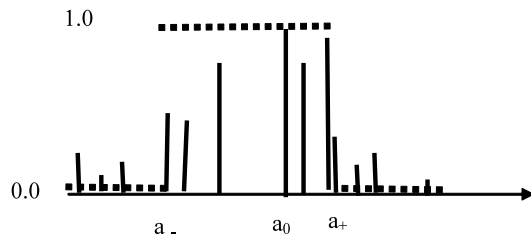


Figure 6. Computing the interval representation of numeric values through the principle of justifiable granularity by elevating and suppressing respective membership grades

6. Higher Order Information Granules

Essentially, there are two fundamental reasons behind the emergence of information granules of higher order: (a) enhancements of interpretability aspects of granular constructs of fuzzy sets, and (b) simplified (and more readable) quantification of diversity of numeric membership degrees. As two examples in the first category, we highlight the concept of shadowed sets and type-2 fuzzy sets^{13,14 15, 16} induced by type-1 fuzzy set. When dealing with the second group of underlying supportive structures, we consider aggregation mechanisms, which bring a collection of type-1 fuzzy sets together by forming membership functions or intervals in the unit interval.

6.1. Interpretability of fuzzy sets: a framework of type-2 fuzzy sets

Membership functions are numeric constructs. Detailed numeric values of membership functions are too precise and in this way difficult to interpret [30]. It is advantageous to come up with quantification, say *low*, *medium*, *high* membership levels. In this sense, rather than talking about ‘a membership degree of 0.75’, we might have a far more descriptive quantification such as “*high* membership”. Figure 7 illustrates the essence of this construct. Crucial to this development are fuzzy sets of linguistic values defined in the $[0,1]$ interval of membership grades.

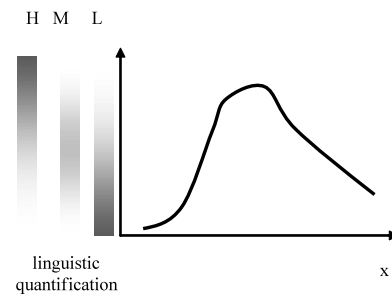


Figure 7. Fuzzy set of type-1 and its linguistic interpretation through fuzzy sets of linguistic quantification

The construction of these fuzzy sets defined in $[0,1]$ and used in the interpretation of the original membership function is required. More formally, let us consider a fuzzy set of interest defined in some space \mathbf{X} , $A: \mathbf{X} \rightarrow [0,1]$. Consider “ r ” fuzzy sets – linguistic descriptors of membership grades $\mathbf{B} = \{B_1, B_2, \dots, B_r\}$, where $B_i: [0,1] \rightarrow [0,1]$. To benefit from the descriptive power supplied by the \mathbf{B} , its fuzzy sets have to be optimized where the optimization could be guided by different design criteria. Here we consider a representation criterion where we request that A is well-defined in terms of the elements of \mathbf{B} . More specifically, the entropy of A expressed in terms of B_i for a certain element x of \mathbf{X} , $H(B_i(A(x)))$. The overall entropy criterion comes in the form

$$V = \int_{\mathbf{X}} \sum_{i=1}^r H(B_i(A(x))) dx. \quad (12)$$

The entropy function $H(z)$ defined in $[0,1]$, $H: [0,1] \rightarrow [0,1]$, satisfies well-known requirements: (a) monotonic increase in $[0, 1/2]$ and monotonic decrease in $[1/2, 1]$ and (b) boundary conditions of the form $H(0) = H(1) = 0$, $H(1/2) = 1.0$. The minimization of (12) is expressed as $\text{Min}_{\mathbf{a}} H(\mathbf{B})$ where \mathbf{a} is a vector of parameters of the fuzzy sets in \mathbf{B} . In particular, when B_i s are taken as triangular fuzzy numbers with $1/2$ overlap between successive fuzzy sets, then the minimum of (12) is achieved by adjusting the modal values of the fuzzy sets (viz. the modal values are the corresponding coordinates of the vector \mathbf{a}). Here mechanisms of evolutionary optimization are worth pursuing. Note that the distribution of the modal values need not be uniform but it rather reflects the characteristics of the data. In case of

finite space \mathbf{X} , the integral in (12) is replaced by the corresponding sum.

$$V = \sum_{j=1}^n \sum_{i=1}^r H(B_i(x_j)). \quad (13)$$

where $n = \text{card}(\mathbf{X})$. Once the fuzzy sets of linguistic quantification of the fuzzy set defined in the finite space have been defined, the original membership function is translated into a string of labels, say LLMMMMHML, etc. with L, M, and H standing for the linguistic labels (say, being associated with B_1 (L), B_2 (M), and B_3 (H)). Note that in this interpretation, any $x \in \mathbf{X}$ and subsequently $A(x)$ invokes (matches) several B_i 's to a nonzero degree. In the formation of the string of the labels, we pick up this B_i to which x belongs to the highest extent. The essence of the linguistic interpretation is illustrated in Figure 8.

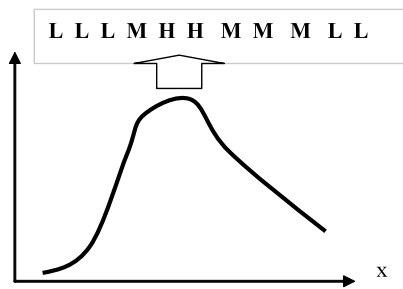


Figure 8. From numeric membership grades to a string of linguistic labels

6.2. Shadowed sets as a three-valued logic representation of fuzzy sets

Membership functions are numeric constructs as far as quantification of membership degrees is concerned. Linguistic quantification enhances interpretability. The concept supported by shadowed sets is developed with the similar goal in mind: rather than dealing with numeric membership degrees, we are concerned with a very limited three-valued quantification characterization scheme of membership functions.

Formally speaking, a shadowed set $A^{17,18,19}$ defined in some space \mathbf{X} is a set-valued mapping coming in the following form

$$A : \mathbf{X} \rightarrow \{0, [0,1], 1\} \quad (14)$$

The co-domain of A consists of three components that is 0, 1, and the unit interval $[0,1]$. They can be treated as degrees of membership of elements to A . These three quantification levels come with an apparent interpretation. All elements for which $A(x)$ assume 1 are called a core of the shadowed set - they embrace all elements that are fully compatible with the concept conveyed by A . The elements of \mathbf{X} for which $A(x)$ attains zero are excluded from A . The elements of \mathbf{X} for which we have assigned the unit interval are completely uncertain - we are not at position to allocate any numeric membership grade. In this region we are faced with a complete uncertainty (don't know quantification). Therefore we allow the usage of the unit interval, which reflects uncertainty meaning that any numeric value could be permitted here. In essence, such element could be excluded (we pick up the lowest possible value from the unit interval), exhibit partial membership (any number within the range from 0 and 1) or could be fully allocated to A . Given this extreme level of uncertainty (nothing is known and all values are allowed), we call these elements shadows and hence the name of the shadowed set. An illustration of the underlying concept of a shadowed set is included in Figure 9.

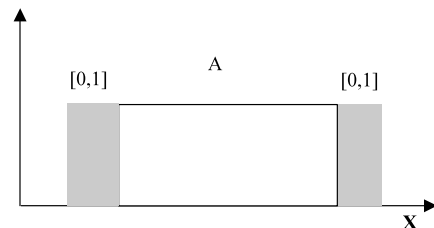


Figure 9. An example of a shadowed set A ; note shadows formed around the cores of the construct.

One can view this mapping (shadowed set) as an example of a three-valued logic as encountered in the classic model introduced by Lukasiewicz. Having this in mind, we can think of shadowed sets as a *symbolic* representation of *numeric* fuzzy sets. Obviously, the elements of co-domain of A could be labeled using symbols (say, certain, shadow, excluded; or a , b , c and alike) endowed with some well-defined semantics. This a conversion of a fuzzy set into a shadowed set is instrumental from the interpretation perspective. For instance, the multimodal membership function illustrated in Figure 10, once the induced shadowed set

has been constructed, comes with a three-valued logic interpretation as succinctly underlined by a string of quantifying values included in the same figure.

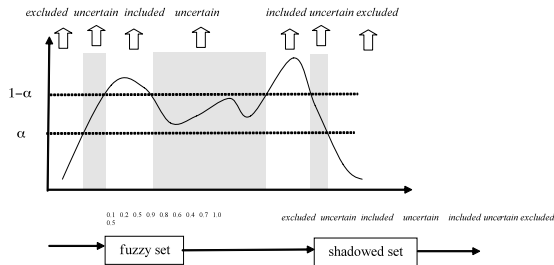


Figure 10. From fuzzy set A to its shadowed set representation and the resulting interpretation.

Construction of shadowed sets

Accepting the point of view that shadowed sets are interpretation vehicles, which are algorithmically implied (induced) by some fuzzy sets, we are interested in the transformation mechanisms translating fuzzy sets into the corresponding shadowed sets. The underlying concept is the one of uncertainty condensation or “localization”. While in fuzzy sets we encounter intermediate membership grades located in-between 0 and 1 and distributed practically across the entire space, in shadowed sets we “localize” the uncertainty effect by building constrained and fairly compact shadows. By doing so we could remove (or better to say, re-distribute) uncertainty from the rest of the universe of discourse by bringing the corresponding low and high membership grades to zero and one and then compensating these changes by allowing for the emergence of uncertainty regions. This transformation could lead to a certain optimization process in which we complete a total balance of uncertainty.

To illustrate this optimization, let us start with a continuous, symmetric, unimodal, and normal membership function A. In this case we can split the problem into two tasks by considering separately the increasing and decreasing portion of the membership function, Figure 11.

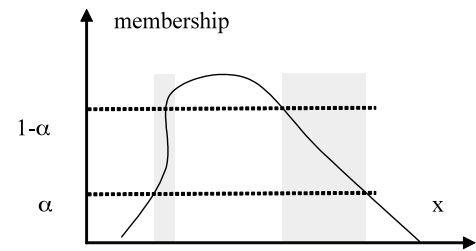


Figure 11. The concept of a shadowed set induced by some fuzzy set; note the range of membership grades (located between α and $1-\alpha$) generating a shadow.

For the increasing portion of the membership function, we reduce low membership grades to zero, elevate high membership grades to one and compensate these changes (which in essence lead to an elimination of partial membership grades) by allowing for a region of the shadow where there are no specific membership values assigned but we admit the entire unit interval as feasible membership grades. Computationally, we form the following balance of uncertainty preservation that could be symbolically expressed as

$$\text{Reduction of membership} + \text{Elevation of membership} = \text{shadow} \quad (15)$$

Again referring to Figure 14, once given the membership grades below α and above $1-\alpha$, $\alpha \in (0, \frac{1}{2})$, we express the components of the above relationship in the form (we assume that all integrals do exist)

Reduction of membership (low membership grades are reduced to zero)

$$\int_{x:A(x) \leq \alpha} A(x) dx. \quad (16)$$

Elevation of membership (high membership grades elevated to 1)

$$\int_{x:A(x) \geq 1-\alpha} (1-A(x)) dx. \quad (17)$$

Shadow

$$\int_{\alpha < A(x) < 1-\alpha} dx. \quad (18)$$

The minimization of the absolute difference

$$V(\alpha) = \left| \int_{x:A(x) \leq \alpha} A(x) dx + \int_{x:A(x) \geq 1-\alpha} (1-A(x)) dx - \int_{\alpha < A(x) < 1-\alpha} dx \right| \quad (19)$$

completed with respect to α is given in the form of the following optimization problem

$$\alpha_{\text{opt}} = \arg \min_{\alpha} V(\alpha) \quad (20)$$

where $\alpha \in (0, 1/2)$. For instance, when dealing with triangular membership function (and it appears that the result does not require the symmetry requirement), the optimal value of α is equal to $\sqrt{2} - 1 \approx 0.4142$. For the parabolic membership functions, the optimization leads to the value of α being equal to 0.405.

For the Gaussian membership function described in the form $A(x) = \exp(-x^2/\sigma^2)$, we get the optimal value of α resulting from the relationship (the calculations here concerns the decreasing part of the membership function defined over $[0, \infty)$)

$$V(\alpha) = \left| \int_0^{\sigma\sqrt{-\ln(1-\alpha)}} (1-A(x)) dx + \int_{\sigma\sqrt{-\ln(\alpha)}}^{\infty} A(x) dx - \int_{\sigma\sqrt{-\ln(1-\alpha)}}^{\sigma\sqrt{-\ln(\alpha)}} dx \right| \quad (21)$$

Let us move on to the most general case in which we do not impose any assumptions as to the form of the membership function. We consider discrete membership values u_1, u_2, \dots, u_N . Denote the minimal and maximal value in this set by u_{\min} and u_{\max} , respectively. The overall reduction of lower membership grades is expressed in the form of the following sum $\sum u_k$ where $\Omega = \{k | u_k \leq \alpha\}$. The elevation of higher membership grades to one leads to the expression $\sum (1-u_k)$ with $\Omega = \{k | u_k \geq u_{\max} - \alpha\}$. For the shadow we consider the cardinality of the set $\Omega = \{k | u_k \in (\alpha, u_{\max} - \alpha)\}$. Then the above conditions translate into the following optimization problem

$$V(\alpha) = \left| \sum_{k \in \Omega} u_k + \sum_{k \in \Phi} (1-u_k) - \text{card}(\Omega) \right|$$

Minimize $V(\alpha)$ with respect to α

(22)

where the range of feasible values of α is given as $[u_{\min}, (u_{\min} + u_{\max})/2]$.

Once optimized, the resulting shadowed set can be treated as a concise descriptor of the corresponding fuzzy set. For the original fuzzy set A , we denote by $\text{core}(A)$, $\text{shadow}(A)$, respectively the core and shadow of the shadowed set induced by A .

The above design process could be generalized in such a way that we introduce a continuous and increasing functional $g(u): [0,1] \rightarrow [0,1]$ that helps quantify the original values of the membership grades when taken into consideration in the balance captured by (19). When reducing membership grades we use the expression

$$\int_{x:A(x) \leq \alpha} g(A(x)) dx \quad (23)$$

while the elevation of membership is guided by the form

$$\int_{x:A(x) \geq 1-\alpha} (1-g(A(x))) dx \quad (24)$$

The typical form of the functional would be a polynomial $g(u) = u^p$, $p > 0$.

7. Statistically Grounded Logic Operators

In fuzzy modeling, logic operators^{20, 21, 22} play a pivotal role in particular when we are concerned with a large number of arguments (which might be the case when dealing with highly dimensional systems). The min and max operators are extreme to a significant extent as their result is an extreme membership degree. In this way the statistical properties of the population of the membership grades involved in the aggregation are not considered at all. T-norms and t-conorms²⁶ lead to the results, which with the increase of the number of arguments quickly converge to 0 or 1. To alleviate these problems, we contemplate incorporating some underlying statistics of the membership degrees which in turn brings us to the idea of statistically driven aggregation – that is statistically grounded logic OR and AND operators²³ (SOR and SAND, for brief)

SOR logic connectives

The (SOR) connective is defined as follows. Denote by $w(u)$ a monotonically non-decreasing weight function from $[0,1]$ to $[0,1]$ with the boundary condition $w(1) = 1$. The result of the aggregation of the membership grades $\mathbf{z} = [z_1, z_2, \dots, z_N]$, denoted by $\text{SOR}(\mathbf{z}; w)$, is obtained as a result of the minimization of the following expression (performance index) Q

$$Q = \sum_{i=1}^N w(z_i) |z_i - y|. \quad (25)$$

where the value of “ y ” minimizing the above expression is taken as the result of the operation $\text{SOR}(\mathbf{z}, w) = y$. Put it differently $\text{SOR}(\mathbf{z}, w) = \arg \min_{y \in [0,1]} \sum_{k=1}^N w(z_k) |z_k - y|$. The weight function “ w ” is used to model a contribution of different membership grades to the result of the aggregation. Several models of the relationships “ w ” are of particular interest; all of them are reflective of the *or* type of aggregation

(a) $w(z)$ assumes a form of a certain step function

$$w(z) = \begin{cases} 1 & \text{if } z \geq z_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (26)$$

where z_{\max} is the maximal value reported in \mathbf{z} . This weight function effectively eliminates all the membership grades but the largest one. For this form of the weight function, we effectively end up with the maximum operator, $\text{SOR}(\mathbf{z}, w) = \max(z_1, z_2, \dots, z_N)$

(b) $w(z)$ is equal identically to 1, $w(z) = 1$. It becomes obvious that the result of the minimization of the following expression

$$\sum_{i=1}^N |z_i - y|. \quad (27)$$

is a median of \mathbf{z} , $\text{median}(\mathbf{z})$. Subsequently $\text{SOR}(\mathbf{z}, w) = \text{median}(\mathbf{z})$. Interestingly, the result of the aggregation is a robust statistics of the membership grades involved in this operation.

We can consider different forms of weight functions. In particular, one could think of an identity function $w(z) = z$. There is an interesting and logically justified alternative which links the weight functions with the

logic operator standing behind the logic operations. In essence, the weight function can be induced by various t-conorms (s-norms) by defining $w(z)$ to be in the form $w(z) = zsz$. In particular, for the maximum operator, we obtain the identity weight function $w(z) = \max(z, z) = z$. For the probabilistic sum, we obtain $w(z) = (z+z-z*z) = 2z(1-z)$. For the Lukasiewicz *or* connective, the weight function comes in the form of some piecewise linear relationship with some saturation region, that is $w(z) = \max(1, z+z) = \max(1, 2z)$.

In general, the weight functions (which are monotonically non-decreasing and satisfy the condition $w(1)=1$) occupy the region of the unit square. For all these weight functions implied by t-conorms, the following inequality holds $\text{median}(\mathbf{z}) \leq \text{SOR}(\mathbf{z}, w) \leq \max(\mathbf{z})$.

SAND logic connectives

The statistically grounded AND (SAND) logic connective is defined in an analogous way as it was proposed in the development of the SOR. Here $w(z)$ denotes a monotonically non-increasing weight function from $[0,1]$ to $[0,1]$ with the boundary condition $w(0)=1$. The result of the aggregation of $\mathbf{z} = [z_1, z_2, \dots, z_N]$, denoted by $\text{SAND}(\mathbf{z}; w)$, is obtained from the minimization of the same expression (25) as introduced before. Thus we produce the logic operator $\text{SAND}(\mathbf{z}, w) = y$ with “ y ” being the solution to the corresponding minimization problem.

As before, we can envision several models of the weight function; all of them are reflective of the *and* type of aggregation

(a) $w(z)$ assumes a form of some step function

$$w(z) = \begin{cases} 1 & \text{if } z \leq z_{\min} \\ 0, & \text{otherwise} \end{cases} \quad (28)$$

where z_{\min} is the minimal value in \mathbf{z} . This weight function eliminates all the membership grades but the smallest one. For this form of the weight function, we effectively end up with the maximum operator, $\text{SAND}(\mathbf{z}, w) = \min(z_1, z_2, \dots, z_N)$

(b) for $w(z)$ being equal identically to 1, $w(z) = 1$, SAND becomes a median, namely $\text{SAND}(\mathbf{z}, w) = \text{med}(\mathbf{z})$.

(c) more generally, the weight function is defined on a basis of some t-norm as follows, $w(z) = 1 - ztz$. Depending upon the specific t-norm, we arrive at different forms of the mapping. For the minimum operator, $w(z) = 1 - \min(z, z) = 1 - z$ which is a complement of “z”. The use of the product operation leads to the expression $w(z) = 1 - z^2$. In the case of the Lukasiewicz and connective, one has $w(z) = 1 - \max(0, z + z - 1) = 1 - \max(0, 2z - 1)$.

Investigating the fundamental properties of the logic connectives, we note that the commutativity and monotonicity properties hold. The boundary condition does not hold when being considered with respect to a single membership grade (which is completely understood given the fact that the operation is expressed by taking into consideration a collection of membership grades). Assuming the t-norm and t-conorm driven format of the weight function (where we have $w(1) = 1$ and $w(0) = 0$ for *or* operators and $w(0) = 1$ and $w(1) = 0$ for *and* operators) we have $SOR(1, w) = 0$, $SAND(0, w) = 0$. The property of associativity does not hold. This is fully justified given that the proposed operators are inherently associated with the overall processing of all membership grades not just individual membership values.

The possibility and necessity measures determined for the two information granules A and B being articulated in the language of SAR and SAND are expressed in the following manner

$$\begin{aligned} \text{Poss}(A, B) &= SOR(z, w); \quad = a_{itb_i} \\ \text{Nec}(A, B) &= SAND(z, w); \quad z_i = a_i s b_i \end{aligned}$$

8. Conclusions

Fuzzy modeling has emerged as a well-established research area with a lot of well-documented pursuits. While fully acknowledging this state-of-the-art situation, we have identified a number of challenges and showed that these may require a substantial revisiting of the current practices and their careful augmentation. Several of the points being highlighted in this study (including the issue of interpretability which goes far beyond the numeric character of interpretation and

stressed its qualitative facet) indicate that more attention needs to be paid to fuzzy sets of higher order. The statistical knowledge could be of interest when dealing with ways of logic aggregation (logic operators) and we have underlined that the approach of this nature offers an interesting alternative to the existing models based predominantly on t-norms and t-conorms.

References

1. F. Wenstøp, Deductive verbal models of organizations, , *Int. Journal of Man-Machine Studies*, **8** (3) (1976) 293-311
2. W. J. M. Kickert, Towards an analysis of linguistic modeling, *Fuzzy Sets and Systems*, **2** (4) (1979) 293-307.
3. R.M. Tong, A control engineering review of fuzzy systems, *Automatica*, **13** (1977) 559-569.
4. W. Pedrycz, Numerical and applicational aspects of fuzzy relational equations, *Fuzzy Sets and Systems*, **11** (1-3) (1983) 1-18.
5. E. Czogala, W. Pedrycz, On identification in fuzzy systems and its applications in control problems, *Fuzzy Sets and Systems* **6**, (1), (1981) 73-83.
6. A. Kandel, W. J. Byatt, Fuzzy processes, *Fuzzy Sets and Systems*, **4** (2) (1980), 117-152.
7. L. Magdalena, What is Soft Computing? Revisiting possible answers, this issue.
8. J. Kacprzyk, S. Zadrozny, Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools, *Information Sciences*, **173** (4) (2005) 281-304.
9. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, N. York, 1981)
10. K. Kiliç, O. Uncu, I. B. Türksen, Comparison of different strategies of utilizing fuzzy clustering in structure identification, *Information Sciences*, **177** (23) (2007) 5153-5162.
11. W. Pedrycz and F. Gomide, *An Introduction to Fuzzy Sets: Analysis and Design* (MIT Press, Cambridge, MA, 1998).
12. I. Ozkan, I.B. Türkşen, N. Canpolat, A currency crisis and its perception with fuzzy C-means, *Information Sciences*, **178** (8) (2008) 1923-1934.
13. J. M. Mendel, Advances in type-2 fuzzy sets and systems, *Information Sciences*, **177**, (1) (2007) 84-110.
14. M.S. Yang, D.C. Lin, On similarity and inclusion measures between type-2 fuzzy sets with an application to clustering, *Computers & Mathematics with Applications*, **57**, (6) (2009) 896-907.

15. D. Wu, J. M. Mendel, A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets, *Information Sciences*, **179** (8) (2009) 1169-1192.
16. C. L. Walker, E. A. Walker, Sets with type-2 operations, *Int. Journal of Approximate Reasoning*, **50** (1) (2009) 63-71.
17. W. Pedrycz, Shadowed sets: representing and processing fuzzy sets, *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, **28**, (1998) 103-109.
18. W. Pedrycz, Interpretation of clusters in the framework of shadowed sets, *Pattern Recognition Letters*, **26** (15) (2005) 2439-2449.
19. W. Pedrycz, Shadowed sets: bridging fuzzy and rough sets, In: *Rough Fuzzy Hybridization. A New Trend in Decision-Making*, S.K. Pal, A. Skowron, (eds.), (Springer Verlag, Singapore, 1999), pp. 179-199.
20. T. Calvo, A. Kolesárová, M. Komorníková, R. Mesiar, Aggregation operators: properties, classes and construction methods, In: *Aggregation Operators: New Trends and Applications* (Physica-Verlag, Heidelberg, 2002), pp. 1-104.
21. V. Torra, Aggregation operators and models, *Fuzzy Sets and Systems*, **156** (2005) 407 - 410.
22. R. Yager, On ordered weighted averaging aggregation operations in multicriteria decision making. *IEEE Trans. on Systems, Man, and Cybernetics*, **18** (1988) 183-190.
23. W. Pedrycz, Statistically grounded logic operators in fuzzy sets, *European Journal of Operational Research*, **193** (2) (2009) 520-529.
24. E. Klement, R. Mesiar and E. Pap, *Triangular Norms*. (Kluwer Academic Publishers, Dordrecht, 2000).