Transactions on Fuzzy Sets and Systems (TFSS)
URL: https://sanad.iau.ir/journal/tfss/
Online ISSN: 2821-0131
Vol.5, No.2, (2026), 260-275
DOI: https://doi.org/10.71602/tfss.2026.1217371

Article Type: Original Research Article
(Journal Pre-proof)

Granular Encoding-Decoding for the Design of Granular Architectures

Witold Pedrycz*

Abstract. Fuzzy constructs (models) interact with numeric entities. This communication is realized through mechanisms of encoding and decoding. Encoding realizes a representation of input data through a collection of information granules and, as such, can be viewed as a nonlinear transformation of the original numeric entity to some internal (granular) format. The decoding is carried out in the opposite direction: the result at the level of information granules is brought back to the numeric entity. This study provides a unified view of the functionalities and design of these mechanisms by studying their components information granules. The optimization concerns a minimization of loss functions guided by criteria of minimal reconstruction error and a retention of semantics of the codebooks (landmarks) encountered in the encoding and decoding procedures. The role of triangular fuzzy sets is discussed along with associated learning mechanisms. Illustrative applications to hierarchical models and fuzzy cognitive maps are covered.

AMS Subject Classification 2020: 60T05
Keywords and Phrases: Encoding-decoding, Embedding, Granular Computing, Reconstruction error, Fuzzy cognitive maps.

1 Introduction

Fuzzy encoding and decoding mechanisms have been commonly used in the technology of fuzzy sets in numerous ways. Quite commonly they are referred to as fuzzification and defuzzification to descriptively allude to the ways in which fuzzy set constructs interact with the external world operating at the level of numeric data. Encoding and decoding utilize collections of fuzzy sets (or more generally, information granules) that could be considered as reference fuzzy sets or a vocabulary (codebook) of semantically sound concepts. Encoding can serve as a preprocessing layer. Decoding acts as a postprocessing layer. The optimization of reference fuzzy sets could be guided by a few objectives including reconstruction error. There are optimization mechanisms that are invoked when guiding the design of fuzzy models, in particular rulebased models and the optimization of reference fuzzy sets is associated with an overall minimized loss function whose optimization involves all components of the models.

It is worth stressing that information granules forming encoders and decoders are designed with the use of various criteria. We identify the main criteria and elaborate on the way in which they are implemented. In general, there are two key directions. The one concentrates on semantic soundness where information granules are characterized by a number of appealing criteria (including coverage and indistinguishability). The other criteria are data oriented where we request that information granules are legitimate (viz. justified)

*Corresponding Author: Witold Pedrycz, Email: wpedrycz@ualberta.ca, ORCID: 0000-0002-9335-9930
Received: 8 September 2025; **Revised**: 16 September 2025; **Accepted**: 16 September 2025; **Available Online**: 25 October 2025; **Published Online**: 7 November 2026.

How to cite: Pedrycz W. Granular encoding-decoding for the design of granular architectures. *Transactions on Fuzzy Sets and Systems.* 2026; 5(2): 260-275. DOI: https://doi.org/10.71602/tfss.2026.1217371

in terms of underlying data. In this study, we revisit them carefully, bring forward some representatives and also develop new ones that tend to simultaneously capture the semantics and data-oriented facet of the construct.

When it comes to semantic soundness, we claim that the family of fuzzy set landmarks defined over the space of reals exhibits several intuitively appealing requirements that exhibit semantic relevance [1] including normality of membership functions, a limited number (adhering to the well-known 7 ± 2 principle), distinguishability (meaning that each fuzzy set retains its meaning), and coverage of the space (to ensure that any numeric input is represented through the fuzzy sets of the codebook).

Encoding and decoding are commonly encountered functional modules in granular or fuzzy constructs, in particular. Their crucial role is to interface the real-world environment of predominantly numeric data within the abstract layer at which the models are positioned and function; refer to Figure 1.

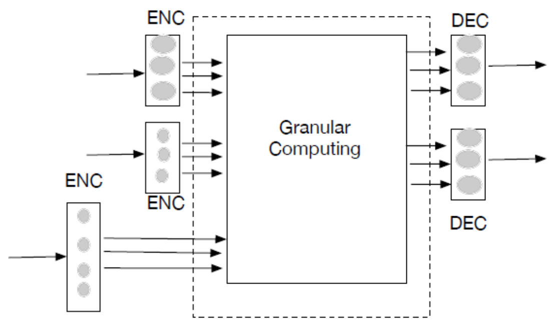

Figure 1: An overall environment of granular architectures constructed in the realm of information granules with the communication mechanisms delivered by encoders (ENC) and decoders (DEC). The dotted line visualizes the delineation between the physical worlds and the world of abstract information granules using which they interact.

Two main structural layers are visibly delineated: the one that concerns data and a way of interaction with the external (physical) world and the abstraction level at which granular constructs are formed and functions communicating their results to the external environment. Encoders and decoders are vividly present in granular architectures, especially fuzzy sets.

Let us recall a few notable classes of models such as Mamdani models, Takagi-Sugeno (TS) models, and logic expressions (built with the aid of logic neurons).

Mamdani fuzzy models. In the generic format, the rules forming the core of this architecture come in the form if $\boldsymbol{x}$ is A_i then y is B_i where A_i and B_i are fuzzy sets. The knowledge base is encapsulated in the form of a fuzzy relation that aggregates the Cartesian products of A_i and B_i whereas the inference mechanisms are guided by relational computing, in particular max-min composition. Evidently, these models operate at the level of information granules so the function of encoders and decoders is critical to the performance of the model at the numeric level as well as a vehicle to provide interpretation capabilities. The information granules A_i constitute an encoder using which any numeric input is encoded. The result is decoded with the decoder equipped with a collection of fuzzy sets B_i subsequently giving rise to the numeric result (e.g., a control action).

TS models. As the rules are in the form if $\boldsymbol{x}$ is A_i then $y = L_i(\boldsymbol{x})$, with L_i standing for a local (usually linear) function, the encoding process is present, however, there is no requirement for the decoding.

Logic models. The models are expressed as a collection of logic expressions $L_1(a, b, c, \ldots), L_2(a, b, c, \ldots), \ldots$ where a, b, c are truth values (in $[0, 1]$) of the logic variables combined by logic connectives (and, or, not, implication, equivalence, etc.) and realized through various t-norms and t-conorms [2]. The construct is positioned at some level of abstraction so the required usage of the encoder becomes evident. The logic models resemble Mamdani models, however, they are more diversified as the logic expressions are flexible by being built with the aid of logic variables and various logic connectives.

The above illustrative examples clearly underline the need for analyzing and designing encoders and decoders as functional modules. Their role is viewed from two essential points of view:

- as a way of endowing granular (including fuzzy) models with interpretability and explainability faculties, and

- as a vehicle supporting a data-oriented design process of building granular models where the existing experimental evidence is taken into consideration.

Interpretability addresses the issues of understanding results; here information granules are supportive to convey the results with some transparent semantics that associates with information granules. Because of the space of information granules involved in the realization of granular models, a suitable criterion guiding the formation of information granules becomes instrumental in the explanation of associations among the granules.

Bearing these two positions in mind, we prudently investigate existing approaches and cast them in the context of interpretability and data-oriented design aspects.

The objective of this study is to revisit existing methods of encoding and decoding and establish a general perspective on the mechanisms of the granulation-degranulation process. In more detail, our main objectives are: (i) identifying a systematic view of the role of information granules in formulating requirements for encoding and decoding, (ii) formulating underlying loss functions to quantify the quality of these two mechanisms along with the pertinent optimization mechanisms including principles of Granular Computing [3]-[6] and reconstruction criterion, (iii) elaboration on the advantages of triangular fuzzy sets in onedimensional encoding and mechanisms of clustering, and (iv) elaborating on the role of granular embedding in the construction of fuzzy architectures including fuzzy cognitive maps and hierarchical modeling.

The unified and holistic view of the problem delivers a substantial level of originality. We argue that encoding-decoding mechanisms play an important role in the accommodation of fuzzy sets into the design of fuzzy constructs and facilitate their incorporation into Machine Learning (ML) and promote synergy between the methodology of fuzzy sets and ML, especially in supporting learning mechanisms and enhancing interpretability. The term embedding could be regarded as a synonym of encoding. From the perspective of fundamentals of fuzzy sets, we demonstrate the emergence of fuzzy sets of order-2 and interval-valued order-2 fuzzy sets.

There is a growing interest in studies on the encoding-decoding triggered by the concept of embedding currently intensively used in numerous architectures of Machine Learning and this becomes a point of interest to capitalize on the principles of Granular Computing and fuzzy sets in the augmentation of semantics of encoding and decoding and their relationships.

In light of the functionality of encoding, it can also be viewed as embedding and from this perspective its role in functioning in systems could be studied. We concentrate on information granules realized as fuzzy sets and intervals, however, the results could be obtained by following the key points cast in other formal settings.

The organization of the exposure of the material proceeds with a presentation of an overall encodingdecoding mechanism (Section 2), which is followed by a formulation of the essence of encoding-decoding mechanism (Section 3). Triangular fuzzy sets forming reference information granules are investigated in Section 4 whereas in Section 5 a multivariable case is studied with a focus on the results of fuzzy clustering. The design of information granules with the principle of justifiable granularity and its variants is discussed in Section 6. Selected applications are presented in Section 7.

2 Encoding and Decoding: An Overall Functional Perspective

Formally, encoding is concerned with a representation of a numeric entity, either one-dimensional or multidimensional, through a collection of information granules $\mathscr{A} = \{A_1, A_2, \ldots, A_c\}$ [7]. In terms of this representation, Enc$(\cdot)$ is described as the mapping Enc : $x \rightarrow y$ or $\boldsymbol{x} \rightarrow y$ where y is described via a collection of granules in $\mathscr{A}$. Enc : R (or $\boldsymbol{R}^n$) $\rightarrow [0,1]^c$. Quite often, the granules in $\mathscr{A}$ are referred to as reference information granules or granular landmarks. The decoding, Dec$(\cdot)$, is a transformation that considers y and through the use of $\mathscr{A}$ brings y to the original space in which $\boldsymbol{x}$ is positioned. Formally, we have Dec : $[0,1]^c \rightarrow R$ (or $\boldsymbol{R}^n$) : $y \rightarrow x$. The mappings are illustrated in Figure 2.

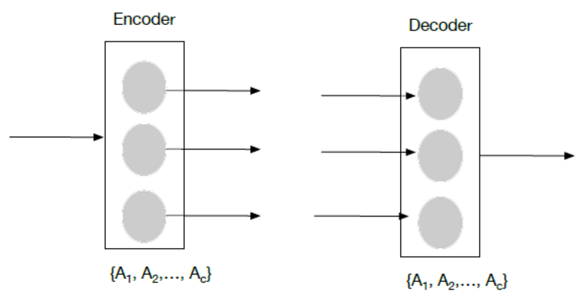

Figure 2: Mappings realized by encoders and decoders

The mechanisms of encoding and decoding are related. In virtue of their nature, we may envision that they are inversely related as the two mappings meaning that Dec(Enc$(x)) = x$ or more descriptively $G^{-1}(G(x)) = x$ where encoding (embedding) is also referred to as a granulation mechanism G and decoding is regarded as degranulation mechanism hence the notation G^{-1}; also G : Enc$(\cdot)$, G^{-1} = Dec$(\cdot)$ The encoding-decoding tandem is illustrated in Fig.3. We also consider that $A_1, A_2, \ldots, A_c$ are associated with their numeric representatives $(m_1, m_2, \ldots, m_c)$ which are of interest when invoking the decoding mechanism.

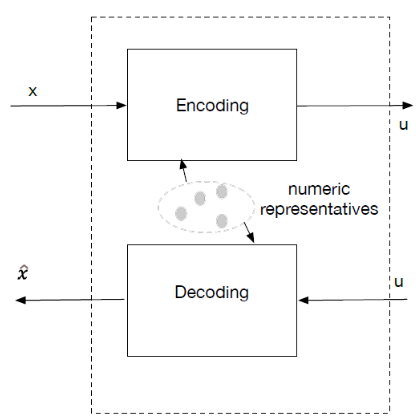

Figure 3: A concept of encoding-decoding modules and the linkages between the modules; $\mathscr{A} = \{A_1, A_2, \ldots, A_c\}$ information granules.

Encoding can be referred to as an embedding mechanism; the terms embedding and encoding are used interchangeably. The embedding leads to the vector $\boldsymbol{y}$ composed of levels of activation (matching) of elements of $\mathscr{A}$ by given x,

$$\boldsymbol{y} = [A_1(x)A_2(x) \ldots A_c(x)] \in [0,1]^c. \tag{1}$$

If $\mathcal{A}$ is composed of sets y is a result of a mechanism of one-hot encoding. For decoding, the computing is carried out in the following form

$$\hat{x} = \sum_{i=1}^{c} A_i(x) m_i. \tag{2}$$

Interestingly, the above expression could be regarded as a realization of an attention mechanism (key-queryvalue) in which $A_i s$ are keys, x stands for a query, and m_i are values - numeric representatives of information granules. It is noticeable that (2) is an example of regression models [8][9].

Ideally, one could anticipate the result of encoding coincides with original x, a suitable criterion sought as a loss function L be regarded as a reconstruction error a distance between x and $\hat{x}$. We express L in the following form

$$L = \sum_{D} \|x - \hat{x}\|^2. \tag{3}$$

Where the sum is taken over some data D from which data xs are being drawn. For intervals forming reference information granules $\mathcal{A}$, the reconstruction error is referred to as a quantization error.

Treating the minimization of the loss function as our objective, a question arises on how to determine the family $\mathcal{A}$ that leads to the minimum of (3). In what follows, we consider one -dimensional and multivariable case with $\mathcal{A}$ described by a collection of fuzzy sets, in particular with triangular membership functions.

3 One-Dimensional Case: Triangular Fuzzy Sets

We assume that $\mathcal{A}$ is formed by triangular fuzzy sets with triangular membership functions with 1/2 overlap present in adjacent fuzzy sets. In other words, triangular fuzzy sets form a partition.

Triangular fuzzy sets have been predominantly used in the technology of fuzzy sets. The popular argument that is brought forward behind their usage was about the simplicity of such membership functions. Nevertheless, the considerations below deliver more compelling arguments along with the quantification of the elements of $\mathcal{A}$. We demonstrate that any triangular fuzzy sets lead to the zero reconstruction error (2).

Theorem 3.1. *Any distribution of triangular fuzzy sets forming a partition with the encoding and deciding described by* (1) *and* (2)*, respectively leads to the zero reconstruction error.*

Proof. For any x, only two information granules are activated, namely $A_i(x), A_{i+1}(x) > 0$. Then the corresponding membership functions are described as $A_i(x) = (\nu_{i+1} - x)/(\nu_{i+1} - \nu_i)$ and $A_{i+1}(x) = (x - \nu_i)/(\nu_{i+1} - \nu_i)$. The decoding result is computed as follows:

$$\hat{x} = [(\nu_{i+1} - x)/(\nu_{i+1} - \nu_i)]\nu_i + [(x - \nu_i)/(\nu_{i+1} - \nu_i)]\nu_{i+1} = x, \tag{4}$$

so the reconstruction result is error-free. $\square$

4 Multivariable Encoding -Decoding Scheme

Now we consider the encoding-decoding scheme for multivariable data in R^n. We commonly use clustering which produces reference information granules. The granules coming in the form of fuzzy sets are generated by Fuzzy C-Means (FCM). For set information granules, one considers K-Means. Given n-dimensional data, as a result of clustering we obtain prototypes $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_c$ using which c-dimensional embedding is completed. FCM produces the result that is in $[0, 1]^c$. For K-Means, the result comes in the form of a one-hot encoding with a single nonzero entry of $u_i(x)$.

Encoding

The result of encoding is expressed in the following form

$$u_i(x) = \frac{1}{\sum\limits_{j=1}^{c} \left(\frac{\|x-\nu_i\|^2}{\|x-\nu_j\|^2} \right)^{1/(m-1)}}, \tag{5}$$

where m stands for the fuzzification coefficient assuming values greater than 1 and $\|\cdot\|$ is the Euclidean distance. The formula (5) is a solution to a certain optimization problem posed below.

Proof. The minimization of (5) in the presence of constraints is transformed to constraint-free minimization completed with the use of Lagrange multipliers. Then we have the augmented objective function in the form $V = \sum\limits_{i=1}^{c} u_i^m \|x-\nu_i\|^2 + \lambda(\sum\limits_{i=1}^{c} u_i(x) - 1)$ where λ is a Lagrange multiplier. After solving the system of equations $dV/du_i = 0$, $dV/d\lambda = 0$ one obtains $mu_i^{m-1}\|\boldsymbol{x}-\boldsymbol{\nu}_i\|^2 + \lambda = 0$ and $\sum\limits_{i=1}^{c} u_i(x) = 1$. Then $(-\lambda/m)^{1/(m-1)} = u_i\|\boldsymbol{x}-\boldsymbol{\nu}_i\|^{-2/(m-1)}$ and $u_i = (-\lambda/m)^{1/(m-1)}\|\boldsymbol{x}-\boldsymbol{\nu}_i\|^{-2}$. Summing up $u_i s$, we have and $u_i\|\boldsymbol{x}-\boldsymbol{\nu}_i\|^{-2/(m-1)} \sum\limits_{i=1}^{c} \|\boldsymbol{x}-\boldsymbol{\nu}_i\|^{-2} = 1$. Thus we obtain (5). $\square$

Decoding

The mechanism of decoding is built through the minimization of a certain loss function. Given the results of encoding and the prototypes, we minimize the following loss function

$$L = \sum_{i=1}^{c} u_i^m \|\hat{x}-\nu_i\|^2. \tag{6}$$

Theorem 4.1. *Given prototypes ν_i, the results of encoding $u_i(x)$ are determined by minimizing (6) where the vector u_i has the entries expressed by (5) and $\|\cdot\|$ is an Euclidean distance. The reconstruction is determined as*

$$\hat{x} = \sum_{i=1}^{c} u_i^m(x)\nu_i / \sum_{i=1}^{c} u_i^m(x). \tag{7}$$

Proof. For the Euclidean distance $\|\cdot\|$, the minimization of (7) is obtained by computing the gradient of L and setting it to zero, $L = \sum\limits_{i=1}^{c} u_i^m \|\hat{x}-\nu_i\|^2 = \sum\limits_{i=1}^{c} u_i^m (\hat{x}-\nu_i)^T(\hat{x}-\nu_i)$ and $\nabla_{\hat{x}} L = 2\sum\limits_{i=1}^{c} u_i^m(\hat{\boldsymbol{x}}-\boldsymbol{\nu}_i) = 0$. $\square$

It is worth noting that in general the result of decoding x does not coincide with x and the resulting discrepancy is referred to as reconstruction (or granulation-degranulation) error.

Both in the one-dimensional and multivariable case, we can regard the result of embedding as an order-2 information granule, viz. the granule defined in the finite space of reference information granules $\mathcal{A}$.

FCM and K-Means generate prototypes. Some other clustering methods such as hierarchical clustering or DBSCAN can be considered however, as these methods do not generate prototypes the prototypes could be determined by further splitting original clusters by involving the K-Means or FCM.

An observation worth making is about the dimensionality of the resulting constructs. In one-dimensional encoding, the increase of dimensionality is visible as one-dimensional element in $\boldsymbol{R}$ gives rise to a c-dimensional vector of membership grades in the $[0,1]^c$ hypercube. For the n-dimensional input, the result is again an element in $[0,1]^c$ however, one can easily encounter a relationship $c \ll n$.

Depending on the dimensionality of the input space, several architectures of the encoding-decoding process could be envisioned, Fig.4.

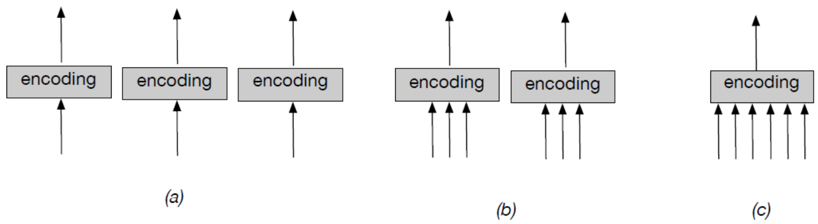

Figure 4: Structures of encoding in multivariable space: (a) Single input encoding and decoding, (b) encoding and decoding completed for groups of inputs variables, (c) multivariable encoding and decoding

Here we design n separate encoder-decoder modules.

(i) Multiple input encoding and decoding. All variables are processed through clustering and a single encoding-decoding scheme is completed. With the high dimensionality of the input space, the quality of clustering could deteriorate (the concentration effect) so an intermediate solution is to cope with the groups of input variables.

(ii) Group inputs encoding and decoding. Some groups of input variables are selected (the clustering is run in subspaces of input variables) and for each group a separate ending-decoding mechanism is realized.

In contrast, if information granules in $\mathcal{A}$ are intervals, the reconstruction error is nonzero; the generated reconstruction error is a well-known quantization error.

5 Design of Information Granules with the Principle of Justifiable Granularity

In one-dimensional case, the triangular fuzzy sets are optimal in the sense of the minimized loss function and zero reconstruction error (3). In terms of the optimized loss function, the families of such fuzzy sets are indistinguishable as all of them generate the zero reconstruction error. There is an infinite number of possible choices. The triangular fuzzy sets can be distributed arbitrarily across the space.

We narrow down this spectrum of possibilities and augment Ais with well-defined semantics by designing a family $\mathcal{A}$ which incorporates some semantics of the information granules. Here the principle of justifiable granularity [5][10] comes into the picture. The concept of information granularity itsef has exhibited a wide range of applications, cf. [11]-[15]. The principle aims at building information granules by making them justifiable in terms of coverage of data by $A_1, A_2, \ldots, A_c$ (so that these granules are legitimized by the data) and at the same time requesting that they exhibit some semantics stating that A_i has some meaning and is distinguishable from other granules in $\mathcal{A}$. From the formal point of view, the coverage and specificity are described in the following way. The coverage of A quantifies how many data points are covered (contained) in information granule and is computed as a sum $1/N \text{ card}\{x_k \mid x_k \in A\}$ for an interval information granule $[a, b]$ and for a fuzzy set information granule with the set of numeric data $\{x_1, x_2, \ldots, x_N\}$. The specificity states how precise (detailed) the information granule is. For the interval, we regard specificity as a decreasing function f of the length (size) of A, $\text{sp}(A) = f(\text{size}(A))$. In the simplest case, f is a linearly decreasing function, $\text{sp}(A) = 1 - (b - a)/\text{range}$ where range serves as a calibration term which is specified as a spread of values of the data. For a fuzzy set A, we use the representation theorem, determine specificity of the corresponding α-cuts of A, A_α and finally specificity becomes an integral of specificities of A_α for corresponding value of A, $\text{sp}(A) = \int_0^1 \text{sp}(A_\alpha)d\alpha$. The coverage is a sum (or integral) of membership grades, $\text{co}\nu(A) = \int A(x)dx$. As both coverage and specificity are in conflict and both of them are to be maximized, the determination of the information granule A is about setting up the parameters of the granule by maximizing the product $\text{co}\nu(A)\text{sp}(A)$. For the interval, we have $a_{\text{opt}}, b_{\text{opt}} = \arg\max_{a,b}[\text{co}\nu(A)\text{sp}(A)]$.

5.1 A Collection of Information Granules for One-Dimensional Data

As we are concerned with a number of information granules, a simultaneous design of these granules forming $\mathscr{A}$ is sought. The coverage criterion is modified to avoid double counting of data to belong to the two adjacent fuzzy sets. For A_i, we introduce the coverage criterion in the following form

$$\text{co}\nu(A_i) = 1/N \sum_k \Big\{ \max \Big(0, A_i(x_k) - \max_{j \neq i} A_j(x_k)\Big)\Big\}, \tag{8}$$

$i = 1, 2, \ldots, c$.

For all information granules in $\mathscr{A}$, we take the following sum over $A_1, A_2, \ldots, A_c$

$$V = 1/N \sum_k \sum_{i=1}^{c} \Big\{ \max \Big(0, A_i(x_k) - \max_{j \neq i} A_j(x_k)\Big)\text{sp}(A_i)\Big\}. \tag{9}$$

The maximization of V is carried out over the modal values of $m_1, m_2, \ldots, m_c$ of the triangular fuzzy sets. As a result, we obtain optimal distribution of the modal values of the fuzzy sets.

5.2 Multivariable Data

The formation of information granules for multivariable data is initialized by building granules through the clustering process. The prototypes produced by the clustering methods are numeric. Around them we build information granules, namely discs, hypercubes, diamonds, etc. the geometry of granules depends on the assumed form of the distance. We make them granular, say discs of the same radius ρ or consider individual radii. A sound argument could be raised about the nature of the prototypes. They are information granules with some radius ρ centered around $\nu_1, \nu_2, \ldots, \nu_c$. In this way we arrive at $V_i = G(x, \nu_i, \rho_i)$, $i = 1, 2, \ldots, c$. For V_i, the coverage is computed as

$$\text{co}\nu(V_i) = 1/N\text{card}\{x_k \mid \|x_k - \nu_i\| \leq \rho\}. \tag{10}$$

If the data are normalized to $[0, 1]$, the specificity is calculated as $\text{sp}(V_i) = 1 - \rho_i$. The optimality of V_i is determined by selecting ε as $\varepsilon_{\text{opt}} = \arg \text{Max}_\varepsilon[\text{co}\nu(V_i)\text{sp}(V_i)]$. In more detail, the coverage is the count of data contained in the information granule belonging to the i-th clusters. If FCM is being sought, the coverage is computed as a sigma count of the data based on the values of the i-th row of the partition matrix.

The decoding in which V_i are used, the result comes in the form of information granule,

$$\hat{X} = \sum_{i=1}^{c} u_i(x)V_i. \tag{11}$$

While the performance of the decoding where numeric results are produced is described with the use of (2), in case of granular results of decoding the quality is evaluated as the product of coverage and specificity, $\text{co}\nu(x, \hat{X})\text{sp}(\hat{X})$. The higher the value of this product, the better the performance of the decoding.

The refinement of the above method could be completed in two different way: (i) by admitting different values of $\rho_1, \rho_2, \ldots, \rho_c$ and (ii) by designing all information granules en block as discussed above.

5.3 Forming Information Granules with Guidance Provided by Auxiliary Variable

The constructs discussed so far dealt with data defined in some input space. In a number of problems, information granules are built in the presence of auxiliary variables, in particular output variable. Along with the optimization criteria of coverage and specificity, we accommodate the requirement stemming from the characteristics of data in the output space that are associated with the data. We require the information granules are made homogeneous with respect to the values of the values of the output variable. In other words, if the information granule A_i is heterogeneous with respect to the output variable, we discount the originally computed coverage which comes from the existence of diversity of output data.

- regression problems. For regression problems we have the continuous output variable whose standard deviation serves as a discount (calibration) factor, σ_{yi}. We modify original coverage formula as follows

$$\mathrm{co}\nu(A_i|y) = \mathrm{co}\nu(A_i)\exp(-\sigma_{yi}). \tag{12}$$

  The higher the standard deviation, the more visible the reduction of the coverage becomes.

- classification problems. One can treat the entropy criterion $h(\omega_1, \omega_2, \ldots, \omega_p)$ serving as a discount factor in the computing of the coverage.

Another design alternative is governed by the cumulative variance of the output data distributed across the information granules. The variance in the output space (output variable) captured by A_i is computed as

$$\sigma_{yi}^2 = \sum_{k=1}^{N} A_i(x_k)(y_k - \overline{m}_i)^2. \tag{13}$$

Where

$$\overline{m}_i = \sum_{k=1}^{N} A_i(x_k)y_k / \sum_{k=1}^{N} A_i(x_k). \tag{14}$$

The optimization of modal values of A_i is completed through the minimization of F.

$$F = \sum_{i=1}^{c} \sigma(A_i), \tag{15}$$

over the modal values $m_2, m_3, \ldots, m_{c-1}$. The lower bound and upper bounds are fixed; $m_1 = \min_k y_k$, $m_c = \max_k y_k$.

In the optimization problems studied in this section, the optimization involves the modal values of the triangular fuzzy sets. In virtue of the nature of these problems, methods of population-based optimization such as genetic algorithms, particle swarm optimization, ant colonies among others are worth pursuing.

6 Mechanisms of Encoding-Decoding Realized in the Presence of Granular Data

So far, we discussed encoding completed for numeric data. The situation is straightforward as one determines the degrees of membership to the individual elements of A. If instead of single $x \in \boldsymbol{R}$ or $\boldsymbol{x} \in \boldsymbol{R}^n$, one has A as an information granule, it is apparent that A cannot be fully described by a single degree of membership. To address this problem, we consider possibility and necessity degrees which describe an extent of overlap and inclusion of A in A_i, respectively

$$\mathrm{Poss}(A, A_i) = \sup_x [A(x)tA_i(x)], \tag{16}$$

$$\mathrm{Nec}(A, A_i) = \inf_x [(1 - A_i(x))sA(x)]. \tag{17}$$

An illustration of these degrees is covered in Fig. 5 where the landmark A_i is an interval in $\boldsymbol{R}$

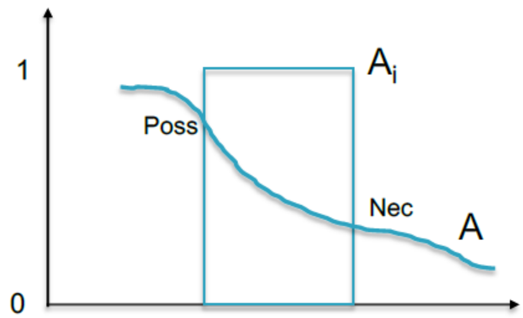

Figure 5: Computing of possibility and necessity degrees for interval A_i and granular input data A

In other words, A is described in terms of A_i by a possibility-necessity pair $(\text{Poss}(A, A_i), \text{Nec}(A, A_i))$. The result of encoding is a $2c$-dimensional vector of membership grades

Enc:

$$A \to [\text{Poss}(A, A_1)\text{Poss}(A, A_2)\ldots\text{Poss}(A, A_c)\text{Nec}(A, A_1)\text{Nec}(A, A_2)\ldots\text{Nec}(A, A_c)]. \tag{18}$$

The result of encoding can be regarded as an interval -valued order-2 fuzzy set.

The decoding can be realized by noting that (16)-(17) form a system of fuzzy relational equations with respect to A. Given $\text{Poss}(A, A_i)$ and A_i, one solves (16) with respect to A. In the same manner, one can view (xx) as fuzzy relational equations with to be solved with respect to A given $\text{Nec}(A, A_i)$ and A_i. Referring to the theory of these relational equations, the largest and the smallest solutions are determined as follows

$$A^{\sim}(x) = A_i(x)\phi\text{Poss}(A, A_i), \tag{19}$$

and

$$A^{*}(x) = A_i(x)\beta\text{Nec}(A, A_i), \tag{20}$$

where $a\phi b = \sup\{c \in [0,1] \mid atc \leq b\}$, $a\beta b = \inf\{c \in [0,1] \mid asc \geq b\}$ with $a, b \in [0,1]$.

As we are concerned with the system of equations, one takes the intersection of the maximal solutions and the intersection of minimal solutions.

7 Selected Granular Architectures with Granular Embedding

Fuzzy embedding is a central component of fuzzy set constructs. We elaborate on the three representative cases, namely fuzzy cognitive maps, hierarchical modeling, and granular non-negative factorization. We show how the inclusion of information granules impacts the architectures and facilitates processing and interpretation of results.

7.1 Fuzzy Cognitive Maps with Granular Embedding of Nodes

Recently graph models have gained a substantial level of interest [16]. Cognitive maps and fuzzy cognitive maps as prediction models have been studied intensively in the realm of fuzzy sets resulting in architectures, learning schemes and applications, cf.[18]. Consider a p-dimensional multivariable time series $z(k)$, $\dim(z) = p$ whose i^{th} coordinate of z is the value of the i^{th} time series reported for the k^{th} time moment, $i = 1, 2, \ldots, p$. Each series is represented by a single node of the map. The weighted linkages between any two nodes (time series) of the graph that capture relationships among the series are displayed as directed edges of the graph.

The values of the weights are confined to the $[-1, 1]$ interval where positive values denote excitatory linkages and negative values quantify the inhibitory associations. The underlying formulas governing the dynamics of the map are expressed as $z_i(k+1) = \phi(z_1(k), z_2(k), \ldots, z_i(k), \ldots, z_p(k), W)$, $i = 1, 2, \ldots, p$. $z_i(k)$ are the activation level of the i-th node (value of the i-th time series) observed in the k-th discrete time moment. Where ϕ is a monotonically increasing function, $\phi : R \to [0, 1]$; a sigmoid function is a frequently considered alternative. W is a matrix of connections (parameters) describing the dynamics of the relationships among the concepts.

There are two key design aspects: (i) the first one is concerned with the representation of numeric time series at the level of abstract entities such as activation levels of the corresponding nodes, and (ii) learning the parameters of the map (W) the problem focuses on the modifications of the connections of the map through the minimization of some loss function. Again, loss function is commonly expressed as a sum of squared errors between the predicted values at each node and the one produced by the map.

Let us note that cognitive maps are abstract constructs developed at a higher level of abstraction than the numeric values of the series. The simple alternative reported in the literature is to normalize the single numeric time series to $[0, 1]$ and view the activation levels. Another interesting option [19] is to explore clustering the time series and view each cluster as a concept (node) of the map. Thus the embedding is carried out through fuzzy clustering. In essence, we cluster many-dimensional data and then associate each cluster (prototype). Therefore, the number of nodes is equal to the number of clusters.

Here we look at the node embedding of the cognitive map which offers a more flexible option to realize mapping from numbers to the granular representation. The numeric values of each time series are encoded by a collection of information granules (triangular fuzzy sets), see Section 3 which in essence leads to embedding of the individual nodes, see Fig. 6. In this way, we have p nodes of the map. At the same time, the node is endowed with the decoder functional module which is essential to realize communication with the external world in terms of forming numeric values of the prediction during the learning of the map.

The embedding process realizes a mapping $z_i(k) : R \to x_i(k) \in [0, 1]^c$. For the simplicity of this study, we consider the same number of information granules for all variables.

The computing is done on the $x_i(k)s$. We have the prediction results in the form

$$\hat{x}_i(k+1) = \phi(x_1(k), x_2(k), \ldots, x_i(k), \ldots, x_p(k), W), \tag{21}$$

$i = 1, 2, \ldots, p$, $\dim(\hat{x}_i) = c$. In terms of coordinates, we describe (15) as follows

$$\hat{x}_{i,t}(k+1) = \phi(\sum_{j=1}^{p} \sum_{s=1}^{c} w_{ij,ts} x_{j,s}(k)), \tag{22}$$

$i = 1, 2, \ldots, p$; $t = 1, 2, \ldots, c$. W is the four-dimensional tensor of dimensionality $p^* p^* c^* c$.

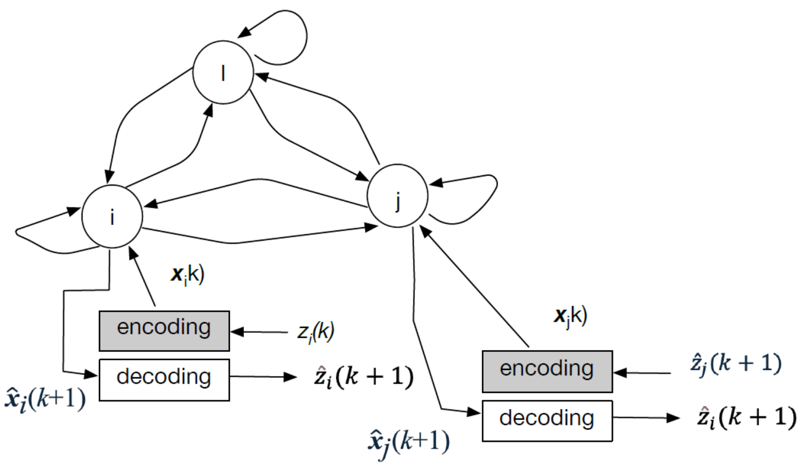

Figure 6: Fuzzy cognitive map with node embedding and decoding (the embedding and decoding are represented as boxes shown next to the nodes of the map).

The learning is carried out in a supervised format given input-output pairs of data $z(k) = [z_1(k)z_2(k)\dots z_p(k)]$, $k = 1, 2, \dots, N$ and $t(k) = [z_1(k+1)z_2(k+1)\dots z_p(k+1)]$, $k = 1, 2, \dots, N-1$. In the learning, two options are considered.

(i) optimization at the level of embedding involving the pairs $(x(k), t\sim(k))$ where $x(k) = \mathrm{Enc}(z(k))$ and $t^{\sim}(k) = \mathrm{Enc}(t(k))$; $\dim(x(k)) = cn$, $\dim(t^{\sim}(k)) = cn$. $\hat{x}_i(k+1)$ is computed following (15). Here the loss function is expressed as

$$L = \sum_{D} \|\hat{x}(k) - t^{\sim}(k)\|^2. \tag{23}$$

(ii) optimization at the level of decoded results. Here as before $x(k) = \mathrm{Enc}(z(k))$. The computed prediction $\hat{x}_i(k+1)$ is decoded, yielding $\hat{z}_i(k+1) = \mathrm{Dec}(\hat{x}_i(k+1))$. Following the decoding process

$$\hat{z}_i(k+1) = \sum_{j=1}^{c} \hat{x}_{ij}(k+1)m_{ij}. \tag{24}$$

With m_{ij} denoting the prototypes of the triangular fuzzy sets, $i = 1, 2, \dots, p$; $j = 1, 2, \dots, c$. The loss function is expressed as

$$L = \sum_{D} \|\hat{z}(k+1) - t(k)\|^2. \tag{25}$$

In the learning process, to enhance interpretability, we keep the values of the connections close to 0, 1, or -1. The values of connections close to $0 \cdot 5$ and $-0 \cdot 5$ are most uncertain and should be avoided (penalized during the learning). The form of the regularization term commonly used in ML learning to capture takes on the entropy-like criterion $h : [-1, 1] \to [0, 1]$ where $h(u) = 4|u|(1 - |u|)$. The loss functions expand (23) and (25) and read as follows

$$L = \sum_{D} \|\hat{x}(k) - t^{\sim}(k)\|^2 + \mu\|W\|, \tag{26}$$

or

$$L = \sum_{D} \|\hat{z}(k+1) - t(k)\|^2 + \mu\|W\|, \tag{27}$$

where μ is a non-negative calibrating coefficient whose value is optimized on the validation set.

B. Hierarchical structure of interpretable models

In multidimensional system modeling engaging fuzzy models, quite often we are concerned with a stepwise development of the models where at each successive level of hierarchy some variables are considered [20]-[23]. This stepwise development of fuzzy models is beneficial to the design itself and the understanding of the model constructed in this well. The stepwise design is both of interest to regression and classification problems. An overall structure is illustrated in Fig. 7. The original variables are embedded following the criteria discussed in Sections 3-4.

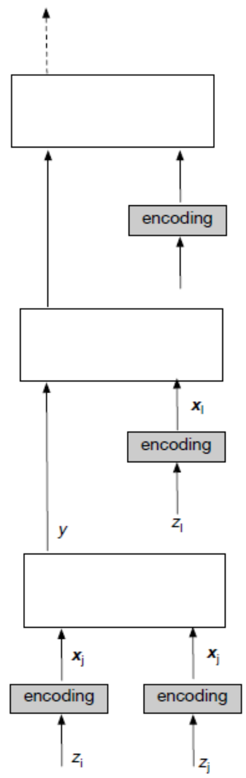

Figure 7: A hierarchical architecture of granular model

The criteria of interest are those in which the output variable is used. Following the embedding the variables are ordered from the perspective of their representation abilities of the input-output relationships. The two variables of the highest value of V are selected first. For them the model is constructed. It could be a neural network, polynomial or a fuzzy logic expression constructed with the aid of t-norms and t-conorms. In essence in this way, we build a fuzzy logic neural network transforming the result of embedding that is an element on the hypercube to the result ' positioned in the $[0,1]$ interval. First, z_i and z_j are encoded yielding $\boldsymbol{x}_i = \mathrm{Enc}(z_i)$, $\boldsymbol{x}_j = \mathrm{Enc}(z_j)$, $\dim(\boldsymbol{x}_i) = n_i$, $\dim(\boldsymbol{x}_j) = n_j$. The two-layer fuzzy neural network with h elements in the hidden layer is described in the form

$$\begin{aligned} z &= \mathrm{AND}(x_i, x_j, G), \\ y &= \mathrm{OR}(z, F). \end{aligned} \tag{28}$$

where G and F are matrices of connections, $G = [G_{ij}]$, $F = [f_{ij}]$. AND and OR are logic composition operators ($t-s$ and $s-t$) using triangular norms. Next we concatenate $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, $\boldsymbol{x}^{\sim} = [\boldsymbol{x}_i\boldsymbol{x}_j]$, $\dim(\boldsymbol{x}^{\sim}) = n_i + n_j$. Element-wise, for the above expressions, we have

$$
\begin{aligned}
z_j &= T_{l=1}^{n_i+n_j}(x_l^{\sim} sG_{jl}), \quad j = 1, 2, \ldots, h, \\
y_j &= S_{l=1}^{h}(z_l tF_{jl}), \quad j = 1, 2, \ldots, h,
\end{aligned}
\tag{29}
$$

where t and s stand for t-norm and t-conorm, respectively.

For the classification problems with r classes, the resulting output of the fuzzy neural network is additionally transformed by the softmax operation meaning that y_j is transformed as $\exp(y_j)/\sum_{l=1}^{h}\exp(y_l)$.

An alternative architecture comprises a three-dimensional fuzzy relation (fuzzy tensor) $G = [G_{ijl}]$ where now we have

$$
z_j = T_{l_1,l_2}^{n_i,n_j}(x_{l1} s x_{l2} G_{jl1l2}), \quad j = 1, 2, \ldots, h. \tag{30}
$$

Next successive layers of the hierarchy are being formed by admitting the variable whose embedding result produces the highest value of the optimization criterion of the product of coverage and specificity.

A hierarchical structure involves only two variables at the first layer of the hierarchy and single variables present at the successive stages. The learning is completed in a supervised mode.

At the first level two variables are embedded by realizing encoding with the aid of c fuzzy sets. Next at the second and successive layers the variable with the highest optimization criterion is added and processed with y coming from the previous layer.

8 Conclusion

The methodology of constructing an encoding-decoding functional module has been formalized and we put forward a systematic design process guided by semantically sound loss functions. The role of these modules in the development of fuzzy constructs has been emphasized and associations with Machine Learning have been identified. The flexibility of embedding implied by the number and distribution of information granules in input space could be explored in the design of deep learning models with fuzzy sets. The emergence of information granules of higher order present in the encoding-decoding mechanisms is emphasized. With the unified perspective offered in this study, some follow up open issues worth pursuing which might explore some promising insights:
(i) a detailed parametric analysis of the codebooks, associated information granules and their impact on Machine Learning architectures regarding enhanced optimization and interpretability, (ii) extensions to other formal environments of Granular Computing, and (iii) exploration of hierarchies of encoding and decoding mechanisms.

Conflict of Interest: The author declares no conflict of interest.

References

[1] Moral JMA, Castiello C, Magdalena L, Mencar C. *Explainable Fuzzy Systems.* Springer; 2021.

[2] Pedrycz W, Rocha AF. Fuzzy-set based models of neurons and knowledge-based networks. *IEEE Transactions on Fuzzy Systems.* 1993; 1(4): 254-266. DOI: https://doi.org/10.1109/91.251926

[3] Zadeh LA. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems.* 1997; 90(2): 111-127. DOI: https://doi.org/10.1016/S0165-0114(97)00077-8

[4] Zadeh LA. From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*. 1999; 46(1): 105-119. DOI: https://doi.org/10.1109/81.739259

[5] Pedrycz W. Granular computing for data analytics: A manifesto of human-centric computing. *IEEE/CAA Journal of Automatica Sinica.* 2018; 5(6): 1025-1034. DOI: https://doi.org/10.1109/JAS.2018.7511213

[6] Qin J, Martnez L, Pedrycz W, Ma X, Liang Y. An overview of granular computing in decision-making: Extensions, applications, and challenges. *Information Fusion.* 2023; 98: 101833. DOI: https://doi.org/10.1016/j.inffus.2023.101833

[7] Zhu X, Pedrycz W, Li Z. Granular encoders and decoders: A study in processing information granules. *IEEE Transactions on Fuzzy Systems.* 2017; 25(5): DOI: https://doi.org/10.1109/TFUZZ.2016.2598366

[8] Nadaraya EA. On estimating regression. *Theory of Probability & Its Applications.* 1964; 9(1): 141-142. DOI: https://doi.org/10.1137/1109020

[9] Watson GS. Smooth regression analysis. *Sankhy: The Indian Journal of Statistics, Series A(1961-2002).* 1964; 26(4): 359-372.

[10] Pedrycz W, Wang X. Designing fuzzy sets with the use of the parametric principle of justifiable granularity. *IEEE Transactions on Fuzzy Systems.* 2016; 24(2): 489-496. DOI: https://doi.org/10.1109/TFUZZ.2015.2453393

[11] Li W, Zhai S, Xu W, Pedrycz W, Qian Y, Ding W, Zhan T. Feature selection approach based on improved fuzzy c-means with principle of refined justifiable granularity. *IEEE Transactions on Fuzzy Systems.* 2023; 31(7): 2112-2126. DOI: https://doi.org/10.1109/TFUZZ.2022.3217377

[12] Shan D, Lu W, Yang J. Interval granular fuzzy models: Concepts and development. *IEEE Access.* 2019; 7: 24140-24153. DOI: https://doi.org/10.1109/ACCESS.2019.2899830

[13] Shi W, Karastoyanova D, Ma Y, Huang Y, Zhang G. Clustering-based granular representation of time series with application to collective anomaly detection. *IEEE Transactions on Instrumentation and Measurement.* 2023; 72: 2530612. DOI: https://doi.org/10.1109/TIM.2023.3325521

[14] Zhang R, Xu K, Zhu S, Xing M, Quan Y. Modeling of number of sources detection under nonideal conditions based on fuzzy information granulation. *IEEE Transactions on Aerospace and Electronic Systems.* 2023; 59(2): 1749-1757. DOI: https://doi.org/10.1109/TAES.2022.3204925

[15] Brcena JLC, Ducange P, Marcelloni F, Renda A. Increasing trust in AI through privacy preservation and model explainability: Federated learning of fuzzy Regression trees. *Information Fusion.* 2025; 113: 102598. DOI: https://doi.org/10.1016/j.inffus.2024.102598

[16] Hu F, Deng Z, Wang G, Xie Z, Choi KS, Wang S. Graph fuzzy system for the whole graph prediction: Concepts, models, and algorithms. *IEEE Transactions on Fuzzy Systems.* 2024; 32(3): 1383-1398. DOI: https://doi.org/10.1109/TFUZZ.2023.3325458

[17] Axelrod R. *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton University Press; 2015.

[18] Kosko B. Fuzzy cognitive maps. *International Journal of Man-Machine Studies.* 1986; 24(1): 65-75. DOI: https://doi.org/10.1016/S0020-7373(86)80040-2

[19] Pedrycz W. The design of cognitive maps: A study in synergy of granular computing and evolutionary optimization. *Expert systems with applications.* 2010; 37(10): 7288-7294. DOI: https://doi.org/10.1016/j.eswa.2010.03.006

[20] Su SF, Chen MC, Hsueh YC. A novel fuzzy modeling structure-decomposed fuzzy system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems.* 2017; 47(8): 2311-2317. DOI: https://doi.org/10.1109/TSMC.2017.2657557

[21] Kamthan S, Singh H. Hierarchical fuzzy logic for multi-input multi-output systems. *IEEE Access.* 2020; 8: 206966-206981. DOI: https://doi.org/10.1109/ACCESS.2020.3037901

[22] Zhang Y, Wang G, Huang X, Ding W. TSK fuzzy system fusion at sensitivity-ensemble-level for imbalanced data classification. *Information Fusion.* 2023; 92: 350-362. DOI: https://doi.org/10.1016/j.inffus.2022.12.014

[23] Zhang Y, Wang G, Zhou T, Huang X, Lam S, Sheng J, Choi KS, Cai J, Ding W. Takagi-Sugeno-Kang fuzzy system fusion: A survey at hierarchical, wide and stacked levels. *Information Fusion.* 2024; 101: 101977. DOI: https://doi.org/10.1016/j.inffus.2023.101977

Witold Pedrycz
Department of Electrical and Computer Engineering
University of Alberta,
Edmonton, AB T6G 2R3, Canada.
E-mail: wpedrycz@ualberta.ca