On Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners: Conception and Validation

Mokhtar Beldjehem
Saint-Ann's University
1589 Walnut Street
Halifax, Nova Scotia, B3H 3S1, Canada
E-mail: mokhtar.beldjehem@usainteanne.ca

Abstract

This paper comprises two parts, the first deals with the conception of a class of Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners, for which a learning scheme was devised that uses an exhaustive search over the fuzzy partitions of involved variables, automatic fuzzy hypotheses generation, formulation and testing, and successive approximation procedure of Min-Max relational equations. The main idea is to start learning from coarse fuzzy partitions of the involved input variables and proceed progressively toward fine-grained partitions until finding the appropriate partitions that fit the data. According to the complexity of the problem at hand, it learns the whole structure of the fuzzy system, i.e. conjointly appropriate fuzzy partitions, appropriate fuzzy production rules, their number and their associated membership functions. The fuzzy relational calculus in the context of approximation of fuzzy relations equations, constitutes a good candidate tool in machine learning, and is especially useful for dealing with inverse problems. The second part deals with verification and validation issues of such learners, validation brings us to a systematic study of value approximation performed during the inference (recall) phase. We provide a rigorous formal mathematical proof that Min-Max rule preserves the property of approximation when it is applied to entities characterized by approximately equal fuzzy values. Hence, using standard Min-Max is a suitable choice in building Hybrid Granular Fuzzy-Neuro or Neuro-Fuzzy Relational Learners, as it is accepted that generalization capability is proportional to value approximation.

Keywords— Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners, Importance, Approximation of Min-Max relational equations, Verification and Validation, Value approximation, Generalization.

1. Introduction, Motivations and Related Work

The concept of a *fuzzy set* was originally proposed by Zadeh in his seminal paper [1] as an extension of the notion of a set by allowing partial membership of an element to a set, and the usage of fuzzy sets theory in relational systems can be traced back early to work by Zadeh [2] who advocated and put the foundations of a theory to model relationships of symptoms and diseases by using the *compositional rule of inference (CRI)* as an inference mechanism. In general, when using a fuzzy system one has the possibility to know how a conclusion has been derived? And why a particular line of reasoning is being followed ? Trace and graphicaly visualize the execution of the CRI used during inference. Due to the parallel firing (execution) of fuzzy rules (no conflict resolution strategy is needed), fuzzy-rule based systems are transparent systems that could provide easily an explanation facilities component that is easy to implement.

Moreover, fuzzy-rule based systems have been proved to be *universal approximators* by Wang and Mendel [3] as multilayer feedforward neural networks by Hornik and al. [4]. Thus, they have high approximation capacity of non-linear functions, and thus they are good candidates to cope with peculiarities of incomplete, imprecise and uncertain knowledge and complexities of real-world problems. Fuzzy rules attempt to capture the "*rules-of-thumb*" approach generally used by experts for decision-making and problem solving. However it is well accepted that crafting manually fuzzy systems to resolve complex large scale real-world problems is a difficult task that is not always obvious for both the designer (the knowledge-engineer, the software engineer) and the domain expert. *Granular and/or soft computing* models as in Zadeh[5-7],Yager and Zadeh[8], Pedrycz[9], Yao[10], Gupta[11], Liu[12], Beldjehem [13-29] combining several paradigms in general and hybrid relational *fuzzy-neuro* models as in Beldjehem [13-29], and/or *neuro-fuzzy* ones as Sinha and Gupta[30] in particular could effectively contribute to the building of next-generation intelligent reliable models. The benefits from adopting such hybridization is its capacity to integrate seamlessly and to account conjointly for both empirical input-output historical data (available measurement data from laboratory test results, simulation results . . .and s.o.) and heuristic domain knowledge that is available from experts, good practices and body of knowledge in general. In addition to resolve the boundary problem, such hybrid fuzzy-neuro and neuro-fuzzy models are transparent, tolerant and could effectively ensure accuracy, performance and interpretability learning trade-offs. The main idea stems from the possibility to use a hybrid fuzzy-neuro or neuro-fuzzy relational system to generate (tune or extract) a knowledge base (KB) or more specifically a rule base (RB) in terms of fuzzy IF-THEN production rules and thus to generate automatically a fuzzy rule-based system by supervised learning from I/O examples (representing instances of the system's behavior).

Most hybrid fuzzy-neuro and neuro-fuzzy models use two, three or four layered-feedforward network and just adapt the *backpropagation algorithm* in order to cope with fuzzy numbers corresponding to the weight of one layer, except few as in Beldjehem [13-29] use novel learning algorithms based on approximation procedure of Min-Max or Max-Min relational equations. Backprop algorithm, which is basically based on a numerical gradient descent method, was initially devised to back-propagate numerical values only, back-propagating fuzzy numbers still a challenge and an open problem, once resolved it will allow us to build true fuzzy multilayered neural networks. Besides the facts that Min and Max are non–differentiable, and the backprop is not suitable for dealing with those models. Moreover the backprog might be stuck in local minima. Following are some useful additional remarks concerning the sate of the art and practice of available Hybrid fuzzy-neuro or neuro-fuzzy models:

- Proliferation of Hybrid fuzzy-neuro or neuro-fuzzy models. How to select the most suitable for the problem at hand is indeed a research problem (It is indeed a multicriteria decision problem!)
- Most use t-norms and t-conorms as operators (to avoid the problem of differentiability in order to use the backprog algorithm) instead of standards Min and Max without justification from the perspective of generalization
- Most use a fixed topology or network architecture that a priori determines the number of rules, which is not always the case in practice
- Their development still is an Art rather than an engineering discipline: Most are crafted in an adhoc manner and are application-dependent
- No systematic design methodologies applied
- No systematic Formal Validation (No proof of correctness), rely on empirical validation only
- Most lack interpretability (a Black box): ANFIS, Fuzzy ART …

On one hand, this has led to intensive research for the conception and development of accurate learning models using various paradigms ranging from statistical Bayesian approach, fuzzy neural networks, fuzzy clustering algorithms, decision trees by conventional ID3 and C4.5 inductive approach of Quinlan [31] to top-down induction of fuzzy decision trees approaches Yuan and Shaw [32]. However to the best of our knowledge, due to (1) the lack of good practices and the unavailability of accurate historical data in general, (2) the incapacity to account conjointly for both historical data and available explicit and heuristic linguistic factual knowledge (3) the lack of interpretability, understandability and explanation facilities; most developed laboratory "proof of concept" prototypes are of low accuracy and thus fail to fulfill their "raison d'être", i.e. helping experts in their practice in regards to decision making for the purposes of problem-solving in the field. On the other hand due to the Black-boxness nature of naïve models in the context of decision making, they still are built on idealizing assumptions such the independence of attributes and due to the lack of probability distribution of possible outcomes they are of limited interest from the decision making viewpoint; by consequent they should be replaced with intelligent soft computational models incorporating explicit formal cause-effect relationships expressed by fuzzy IF-THEN rules that are built or rather machine learned automatically using a data-driven model-free approximation through the hybrid granular soft computing methodology.

The Validation and Verification (V&V) of hybrid granular soft computing models becomes of increasing concern as these systems are fielded and embedded in the every day operations of medical diagnosis, pattern recognition, fuzzy control and other industries--particularly so when life-critical and environment-critical aspects are involved. Two options are possible to validate granular soft computing models, empirical validation or formal validation. One has to distinguish and contrast, between empirical validations such as those based on cross-validation which is merely a statistical method devised to assess the accuracy of a classifier, and a formal systematic ones which are established to study and asses the generalization capacity of a model mathematically. The second option is adopted herein to validate our Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners. Besides integrating non-linearities directly from the learning examples (training set), the additional advantages of such Learners is the inherited property of *value approximation* and *robustness* which is of paramount importance in exhibiting the desired generalizations necessary to process unseen situations (including testing set and validation set).

2. Our Granular Min-Max Fuzzy-Neuro Relational Learner

2.1. Motivations for our learning methodology

We want to conduct herein a case study in connection with problem solving that the human experts are faced with in their practice. In making decisions, experts usually rely on their experiences and intuition rather than applying precise step-by-step algorithms. It is well known that decision making is a very complex and a difficult task to program algorithmically or even to build directly by means of a conventional knowledge-based system or expert systems. Our current focus is to build automatically by learning a system capable to assist an expert in decision making. To this end, we propose to adapt and use a granular hybrid Min-Max fuzzy-neuro relational model for automatic building of an If-Then Fuzzy rule system. More specifically we will investigate the possibility to learn and understand causal relationships between some input variables, and the output variables. Moreover, we will propose the interpretation of the results in terms of weighted fuzzy IF-THEN production rules and the relative importance of the variation of the inputs in relation with the output. In our modeling we use *fuzzy variables* that are *linguistic variable* defined over term sets (or labels of fuzzy sets) and herein represented in terms of trapezoidal membership functions (MFs) and/or possibility distributions for convenience and programmability. When the possible values for a fuzzy variable are symbolic rather numeric, approximations can be represented in terms of a fuzzy set with a corresponding trapezoidal *membership function* (MF). In our modeling we use Fuzzy systems[33] and Zadeh's *possibility theory* [34-35] and more specifically *possibility/necessity measures* which enables us to accurately estimate how much it is possible that we infer a given conclusion, and how much it is necessary that that we infer a given conclusion. We believe that our approach will definitely open the door for next generation intelligent fuzzy systems. Besides machine learning the causal relationships between the inputs and the outputs, they allow the detection of the importance or relevance of each input to a given output which is of paramount importance for an empirical approach of studying and understanding causal relationships aspects and hence providing justification facilities for the validation issues and during the deployment and operation in the field. Thereby, enabling understanding of relative importance of each input and its influence in causing the appearance of a given output.

Zadeh's fuzzy sets [1] and fuzzy logic [33-35] may be considered as a basis for knowledge and meaning representation and is particularly suited for dealing with natural language and medical knowledge peculiarities. We believe that it is the concept of possibility/necessity distributions Zadeh [34-35], rather than the truth, that will play the primary role in manipulating such knowledge for the perspective of drawing conclusions. Possibility theory as originally proposed by Zadeh [34-35], and studied and presented by Yager [36], Dubois and Prade [37], and Olaf [38] provides a formal framework for representing and dealing with ignorance, and uncertainties prevalent in modeling real world problems in a flexible computerized manner straightforwardly. It allows handling uncertainty in a rather coherent qualitative way. Two measures of uncertainty called *possibility* and *necessity* are associated with a possibility distribution. These measures turn out to be a convenient tool for modeling of uncertainty, which allows for the representation of imprecise pieces of information, gradual properties, flexible constraints (expressing preferences), incomplete state of information or partial states of ignorance. However it is well accepted that crafting manually fuzzy systems to resolve complex large scale real-world problems is a difficult task that is not always obvious for both the designer (the knowledge-engineer) and the domain expert. This is due partly to the *cognitive limits* of the human being Miller[39], but also to the difficulty of understanding the intricacies of dimensionality and inherent complexities and peculiarities of large scale real world problems, and in particular when dealing with complex large scale systems. Not to mention the lack of precision in the human-human interaction and communication that affects significantly the knowledge acquisition process during the tandem knowledge-engineer/domain expert relationship. Furthermore once it is undertaken it is labor-intensive, costly, error prone, time-consuming, and done on a trial-and-error basis in an adhoc manner and hence need to be totally or partly automated. This is known as the knowledge acquisition bottleneck problem or the *Feigenbaum bottleneck* and is a common problem for all AI approaches. Soft computing as an automated knowledge acquisition methodology aims at remedying such a problem among others, while substantially reducing the overall development cycle, and improving time to market. Clearly, soft computing provides better alternative solutions to address and resolve such problems.

Various soft computing (SC) techniques have been used to tackle this learning problem from various points of views. However they are based on some idealizing assumptions and no one adopts a holistic approach to resolve such a problem globally, i.e., finding conjointly appropriate fuzzy partitions, fine tuning the membership functions of the labels used in the rules as well as identifying the structure of the fuzzy system (both the required number of rules and rules themselves explicitly) simultaneously. In practice the required number of rules of the system is not known in advance. Indeed learning fuzzy if-then rules is a difficult multi-parameter optimization problem! We have previously devised, developed, empirically validated and deployed a granular hybrid fuzzy-neuro system called Fennec Beldjehem [13-29] that was successfully applied to a difficult problem of biomedical diagnosis [13-17], to image processing and vision engineering [24], to a complex handwriting pattern recognition problem [27], as well as in software quality prediction [26, 29]. Based on our previous work, we first review our model, and we describe this multifaceted problem and then we propose herein an integrated framework to modify the model, accommodate it and extend its ability and scope of applicability for dealing with classification problems by integrating some useful concepts from the human cognitive processes and adding some interesting granular functionalities and knowledge of the domain. In general, the knowledge available is generally heuristic in nature and in making decision experts tend to think on terms of fuzzy rules heuristically.

The basic idea underlying our framework stems from the following interesting remarks about human cognition: Let us first focus our attention on the human problem solving process. In solving problems the human starts from a *coarse description* but if needed iterates and goes gradually to a *fine-grained description* or in-depth details enabling more understanding of the underlying problem until reaching a point where one can effectively find a solution and so stops and does not need any more details. At this point, an excess of precision is not needed (is not necessary) because a certain satisfying trade-offs between

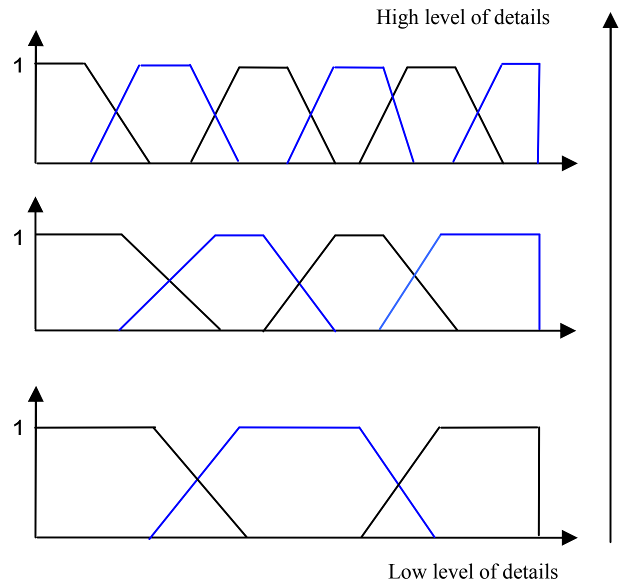

Figure 1: Learning From a Coarse Fuzzy Partition to a Fine-Grained Fuzzy Partition

Precision (*level of details*) and generality of description has been reached and is sufficient and enough for finding a satisfactory approximate solution to the specified problem. Thus after each iteration (increment) a gain of information is obtained enabling more in-depth and more understanding of the underlying situation. Thus, the human converges to a solution gradually by leveraging the level of details. See Figure 1 for more details in connections with a *Granular Soft Computing (GrSC)* setting. Low levels of details allow coarse or general descriptions reflecting crude approximations whereas high levels of details allow specific descriptions reflecting more or less relatively precise approximations (crisps at the extreme). It is appealing and convenient to mimic mechanically or to emulate computationally such a cognitive process in order to automatically build faithfully by learning an appropriate "*good*" fuzzy diagnosis system that exhibits both a high accuracy and a good performance for any problem at hand. This motivates us in building a learning system able to use such abstraction, graduation and granulation mechanisms in a fashion that is akin to the way humans achieve problem solving process. In general the required levels of details necessary in describing rules as well as the required number of rules for solving a problem depends to the degree of complexity of the problem at hand and are unknown and hence we propose to detect and determine them by learning within our framework.

A second aspect of developing information processing systems is the restriction imposed on us by *Millers's Law,* which applies to fuzzy rule-based systems as well. Georges Miller showed that, at any one time, we human being are capable of concentrating on only approximately seven chunks (unit of information, or granules) [39]. One way we humans handle this restriction on the amount of information we can handle at one time is to use *stepwise refinement*. That is, we concentrate first on those aspects that are currently the most important and postpone until later those aspects that are currently less critical. This means, that every aspect is eventually handled, but in order of current importance.

The rational behind using levels of granularity in our framework is obvious for the reader. In addition, this is in the spirit of Zadeh's fuzzy sets and information granulation mechanisms [6, 40].

2.2. The statement of the learning problem

2.2.1 Modeling trough Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners

This framework will be described. Moreover, we will propose the interpretation of the results in terms of weighted fuzzy IF-THEN production rules and the relative importance of the variation of the inputs in relation with the output. In our modeling we use *fuzzy variables* that are *linguistic variables* defined over term sets (or labels of fuzzy sets) and represented in terms of trapezoidal membership functions (MFs) or possibility distributions.

Referring to figure 2, for the sake of simplicity and programmability, this fuzzy set (or equivalently MF) is modeled in term of a couple of crisp sets (or characteristic functions), explicitly an lower envelope (which corresponds to the kernel) and an upper envelope (which corresponds to the support), this representation allows more flexibility in handling and coping with several adjustment operations required during learning as any MF is entirely determined by its two envelopes constituents.

In our modeling we use Zadeh's linguistic variables [42] in addition to possibility theory and more specifically possibility/necessity *measures* which enables us to accurately estimate how much it is possible that a class is stable (or instable), and how much it is necessary that a class is stable (or instable). Besides learning the causal relationships between the inputs and the outputs, they allow the detection of the importance and/or relevance of each input in relation to an output which is of paramount importance for an empirical approach of studying for understanding aspects of the problem under study and hence providing justification facilities for the validation issues. This enables the understanding of *relative importance* of each input and its influence in inferring outputs. Ultimately, this enables us to determine the *minimal subset of most relevant inputs* allowing predicting the possible output accurately. For instance, in a medical diagnosis, the inputs neurons might be certain number of symptoms (say, corresponding to five proteins), our aim is at assigning the appropriate diagnoses (outputs) to a patient among a certain number of possible diagnoses (say, eleven possible diagnoses). Say, eleven output neurons are required in order to represent the eleven possible diagnoses.

We assume herein for convenience that a linguistic variable might have linguistic values represented by labels of fuzzy sets (such as NORMAL, DECREASED, SLIGHTLY DECREASED, INCREASED, VERY INCRESEAD, and s.o.) and interpreted by MFs as illustrated in Figure 3.

For instance, "X_1 is SLIGHTLY INCREASED" indicates a *soft constraint* on possible values rather than a precise characterization of the numerical value to be assigned as it is the case in crisp numerical intervals. Thus in our present modeling, each variable (either input or output) are associated and interpreted by a *fuzzy partition* or equivalently a fuzzy sequence comprising a certain number of terms or granules depending on the level of detail required to describe the variable under scrutiny, the number of which as well as the slopes of membership functions will be determined once a learning session by the *Hybrid Granular Min-Max Fuzzy-Neuro Network* is completed.

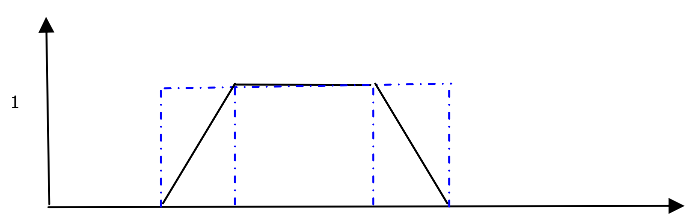

Figure 2: an Unknown Trapezoidal MF as Being Modeled in Terms of Two Envelopes, Lower and Upper Envelope Approximators

2.2.2 Description of the Learning Process and Design issues

The learning is parametric as well as structural. It has to deal with the complexity of the problem and to discover appropriate knowledge chunks, and approximation heuristics for the problem at hand.

Taking into account the degree of complexity of the problem at hand as well as the empirical knowledge contained in the training set, the learning subsystem:

- Start with fuzzy partitions composed of a certain number of granules (say, with c=2), each of which is being modeled in terms of two envelopes, lower and upper envelope approximators. Once, learning is completed a granule will be learned and modeled in term of a membership function (MF) only.
- Identify explicitly the appropriate fuzzy partition for each variable by learning. They are used only as references to generate fuzzy hypotheses. For each variable the appropriate number of granules and the slopes of which will be determined during learning. This information could be either kept or thrown away once the learning is completed without loss of information for the system. As they constitutes only means for generating appropriate membership functions of fuzzy rules and are not used during inference.
- Find the appropriate membership functions for both the antecedents and consequents of every potential rule that is needed to model the problem at hand.
- Ultimately, build the appropriate "good" collection of if-then fuzzy rules (the rule base or knowledge base that consists of a set of linguistic rules), that fits "best" the data that consists of I/O pairs of the training set.

In order to build an automatic workable computational multi-pass learning model some design assumptions are made:

- At each cycle for each input variable X_i the system generates dynamically a fuzzy partition of c granules (starting with c=2, and incrementing c by 1 or 2 at each cycle until reaching a satisfying point). This point constitutes the stopping criterion of our learning mechanism and it reflects too the accuracy level required for the system. It is worth mentioning that increasing c alone does not affect the algorithmic computational complexity of the learning process! It is the number of input variables (n) of the system when it is very large that affects it significantly. We assume to have a reasonable value for n which is almost the case in most classes of real world problems applications.

An output variable may be deal with as an input one, but for the sake of simplicity and programmability we assume that are discrete for our current case study. As the domain expert is more faced with the difficult problem of capturing relationships between the combinations of input variables in relation with a given output variable. In general, for a given output variable the actions (or classes) are well categorized (the number and names of discreet values or granules are known) by the domain expert even thought the slopes of associated MFs have to be questioned during learning.

2.3. Formulation of the Learning Problem

2.3.1 Hypothesis Generation, Formulation and Testing

During learning-time, only one operator is needed to create a fuzzy partition having the required known granularity c. It is the repartitioning operator. It consists to divide dynamically during learning-time the universe of discourse into c overlapping granules. It works from scratch, i.e., there is no need for splitting, or fusion or expanding. A partition is used as reference only and its granules do not necessarily constitute MFs for actual rules as they are only used for formulation of initial fuzzy hypotheses during the generation by the systematic exhaustive search algorithm and they are both scale-dependents and context-dependents. We have no other assumption about the fuzzy partition and we are not interested to argue in such matters like "good" partition. The learning will be done at the rule level rather than at the partition level and hence learning a "good" rule is indeed a crucial issue of utmost importance. A fuzzy partition or covering, also known as a *frame of cognition* is illustrated in Figure 3 (observe how the rightmost and the leftmost granules are shaped); it is a parameterized family (sequence) of membership functions that cover the universe of discourse for every variable either input or output.

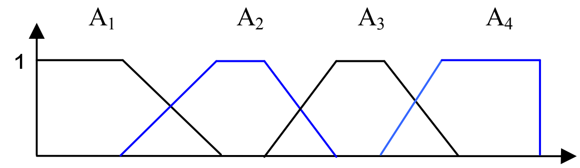

Figure 3: A Fuzzy Partition of Granularity c=4 that is a Superposition of Two Wave Functions Representing an input variable

It is created dynamically by the execution of the repartitioning operator of granularity equals to c during learning-time. In fact, it is obtained by superposition of two wave functions defined over the same universe of discourse X ranging in the interval [a_{min}, a_{max}]. Thus, it is straightforward to extract parameters of granules (MFs) from a given fuzzy partition, as each granule may be considered as an indexed term of the family (or sequence). Moreover the coverage of the universe of discourse (domain) is ensured automatically by construction during learning.

When the possible values for a variable are symbolic rather than numeric, which is almost the case in real-world problems; linguistic approximations can be learned and represented in terms of labels of fuzzy sets with a corresponding membership function. A computationally more efficient and convenient way to characterize it is to use a parametric representation of the MFs of its constituents (called fuzzy members). A fuzzy partition might be thought of as a sequence of granules, each of which is represented by an indexed term. In general as illustrated in Figure 3, kernels are mutually disjoints and every value x of the universe of discourse corresponds to at most two granules. Such desirable property has to be preserved during learning. A_1, A_2 . . . A_i . . . A_c are just synthetic linguistic labels interpreted by fuzzy sets of normalized MFs. A fuzzy partition might be thought of as a synthetic alphabet that the system create by learning for future hypotheses generation. Thanks to this flexible scale-dependent representation, regardless the range of the universe of discourse of an input variable, the terms of the fuzzy partition sequence are explicitly expressed straightforwardly as follows for a trapezoidal MF:

$$\mu_{Ai}(x) = \begin{cases} (x-a)/(b-a) & if \quad a \le x \le b \\ 1 & if \quad b < x < c \\ (d-x)/(d-c) & if \quad c \le x \le d \\ 0 & otherwise \end{cases}$$

When b=c, a trapezoidal shape will degenerate into a triangular shape expressed explicitly as follows:

$$\mu_{Aj}(x) = \begin{cases} (x-a)/(b-a) & if \quad a \le x \le b \\ (d-x)/(d-c) & if \quad c < x \le d \\ 0 & otherwise \end{cases}$$

2.3.2 Learning by Hybrid Min-Max Fuzzy-Neuro Network and Conception Issues

Fuzzy (weighted) rules have been proposed originally by Zadeh [33-35] and have been advocated, used, studied and interpreted by many authors Cayrol et al. [43], Dubois et al. [44] Beldjehem[13-29], Yager[45] and originally machine learned automatically by Beldjehem [13-18]. We will focus in dealing with a multi-input single-output (MISO) system as any multiple-input multiple-output (MIMO) system could be converted to a certain number of MISO systems. Let us start with a model overview: As in Beldjehem [13], we consider herein to design fuzzy-neural possibilistic network architecture according to the scheme Fuzzy to Neural (or to switch from fuzzy systems to neural networks). We use fuzzy if-then weighted rules that are herein of the classification type as in [13-18] and such a rule looks like:

IF (X_1 is w_{k1}, c_{k1}) and (X_2 is w_{k2}, c_{k2}) and (X_3 is w_{k3}, c_{k3}) and (X_5 is w_{k5}, c_{k6}) **THEN** D is D_k

c_{kj} is a weight that represents the grade of importance of " X_j is w_{kj} " in relation with the output D_k. Thus, conversely the weight a_{kj} =1- c_{kj} represents the grade of unimportance of "Xj is w_{kj}"in relation with the same output D_k.

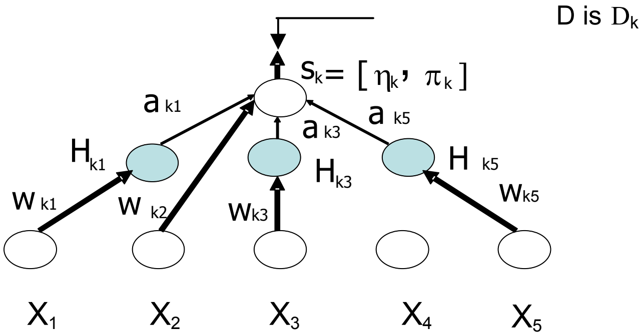

Figure 4: Schematic Representation of the Hybrid Fuzzy-Neuro Possibilistic Min-Max Model Used.

Referring to Figure 4, we propose herein a feed-forward fuzzy-neural possibilistic network. We begin with a brief description of the model: two types of weights are associated with the connections.

Type 1: Direct connections between input cells (X_j) and output cell (s_k) with only synthetic linguistic weights (w_{kj}), interpreted as labels of fuzzy sets, characterizing the variations of the input cells ("X_j is w_{kj} ") with the output cell (s_k), in this case we have a_{kj}= [0,0]=0. Thus ($\prod(X_j;w_{kj}) \vee 0) = \prod(X_j;w_{kj})$. Thus the connection between a hidden cell and output cell simply disappears from the graph allowing direct connection.

Type 2: Connections between input cells (X_j) and output cells (s_k) via intermediate cells (H_{kj}), weights associated to connections between input cells (Xj) and intermediate cells (H_{kj}), are herein artificial or synthetic linguistic (w_{kj}), weights associated to connections between intermediate cells (H_{kj}), and output cell (s_k) are herein numerical intervals ($a_{kj} \subseteq$([0,1]), instead of a scalar value ranging in the interval [0,1]) ($a_{kj} \in$ [0,1]).

w_{kj} are unknown artificial or *synthetic linguistic weights* and a_{kj} are unknown confidence interval that reflects a domain of possible values of unimportance for the corresponding connections. Thus providing much more flexibility for the network.

A learning session starts with a "*blank*" fully connected hybrid fuzzy-neuro network without a priori information concerning the weights, i.e. the weights might be thought of as "*placeholders*" only. Learning is parametric as well as structural. Let us consider now cell activation for an arbitrary output cell (s_k), as illustrated in Figure 3, where only connections used in activation of s_k appear. From the semantic point of view, such a figure reflects a neural representation of an if-then fuzzy weighted rule of control type. Let $\prod(X_j; w_{kj})$ = Sup [$w_{kj} \cap X_j$] be possibility measure associated to fuzzy sets w_{kj} and X_j. And let $N(X_j; w_{kj})$ = Inf [$w_{kj} \cap$ Not X_j] be necessity measure associated to fuzzy sets w_{kj} and X_j . In general our model is governed by the three following abstract fuzzy approximate equations as shown below. Thus allowing the manipulation of fuzzy I/O examples and enabling *approximate learning* reflecting *soft mapping*, this in fact is a departure from conventional learning algorithms.

$$\pi_k = \bigwedge_{j\in\{1,2,3,5\}} (\prod(X_j, w_{kj}) \vee a_{kj}) \tag{1}$$

$$\eta_k = \bigwedge_{j\in\{1,2,3,5\}} (N(X_j, w_{kj}) \vee a_{kj}) \tag{2}$$

$$s_k = [\eta_k, \pi_k] \tag{3}$$

Obviously, each output variable will be assigned an interval as illustrated in equation 3; the inputs of the fuzzy-neuro networks represent certain number of inputs' measurement data of involved input variables and certain number of output variables in terms of possibility/necessity measures. The interpretation by the means of *linguistic approximations* of the output illustrated in TABLE I. The process of linguistic approximation consists of finding a label whose meaning is the same or the closest (according to some metric) to the meaning of unlabeled MF (representing either a fuzzy set or an interval) generated by some computational model (learning in our current study).

Beyond biomedical diagnosis, we are more interested herein by building a class of software tools that justifies and explains its reasoning so that the knowledge and problem solving process is remembered and mimicked by the practitioner in order to tackle the validation and understanding issues. Simply put a system which not only solve the problem of the biomedical diagnosis but also is able to construct a transparent model for the physician, clinicians and biologists (the human problem solver) towards understanding the problem under investigation.

Table 1. The linguistic approximations of certainty values

Certainty value S_k	Linguistic approximation
[0, 0]	IMPOSSIBLE
[0, 0.05]	ALMOST IMPOSSIBLE
[0, 0.1]	SLIGHTLY IMPOSSIBLE
[0, 0.65]	MODERATELY POSSIBLE
[0, 1]	POSSIBLE
[0.35, 1]	QUITE POSSIBLE
[0.9, 1]	VERY POSSIBLE
[0.95, 1]	ALMOST SURE
[1, 1]	SURE

Observe that Maximum ($\vee$) limits lower amplitudes of inputs, we have ($\prod(X_j;\, wkj) \vee a_{kj}) = a_{kj}$ if $\prod(X_j;\, wkj) \le a_{kj}$, and amplifies higher ones ($\prod(X_j;\, wkj) \vee a_{kj}) = \prod(X_j;\, wkj)$, if $\prod(X_j;\, wkj) \ge a_{kj}$, so the Min-Max composition indicates a somewhat excitatory character. It is worthwhile to notice that Min-Max composition as containing Min and Max operations is strongly nonlinear. Furthermore, such model has been formally validated and it has been shown recently (Beldjehem 2006, 2008) that Min-Max composition preserves the value approximation property. Observe that when $a_{kj} = 1$, the term $\prod(X_j;\, wkj) \vee a_{kj}$ (respectively $N(X_j;w_{kj}) \vee a_{kj}$) is deleted in the application of Minimum ($\wedge$). Thus, ensuring the interpretability and transparency of the model. It is now clear that a_{kj} reflects a notion of relative unimportance, we point out herein that it is strongly hard if not impossible to make values assignment to grades of unimportance in practical applications, we will propose a mechanism to learn such grades of unimportance. Semantically, missing edge reflects the non-influence of the input the appearance or presence of the output (or the class). This enables us to determine the minimal subset of the most relevant inputs allowing to infer the output or to identify the outcome, in a fast, transparent, accurate and faithful fashion. Thus, enabling the understanding of the intricacies and complicated issues of the problem under study. Moreover, this permits powerful systems to b implemented on low-cost standards hardware, such as single-chip microcontrollers.

Strengths, benefits and advantages of this Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners include:

- To benefit from the synergism of fuzzy systems and neural networks (are complementary rather than competitive to insure capability enhancement)
- To benefit from the advantages of each one and avoid the weaknesses of each one, to cope with the complexity of multifaceted real-world problems
- To handle a mixture of nominal, ordinal, interval and fuzzy scale-based data in a rather unifying manner within the same conceptual framework
- To start from low-level raw data (training set) to squeeze or grasp high-level knowledge-based computational models for problem-solving. To obtain clarification and to allow verification and validation (V&V)
- To overcome the problem of knowledge engineering known as the knowledge acquisition bottleneck or well-known as the Feigenbaum bottleneck, simplify design complexity
- To learn a high level of representation (IF/THEN fuzzy production rules)
- To work at a higher level of abstraction originating from learning and during inference (or recall)
- To ensure good performance-interpretability tradeoffs
- To ensure robustness, noise tolerance and graceful degradation,
- To build biologically and/or cognitively motivated models
- To understand human perception, cognition and intelligence

3. Resolution of the Learning problem

3.1 *The Learning Algorithm and Implementation Issues*

During a learning session the same learning algorithm is used for each output variable Y_j. Let us briefly describe the learning algorithm that is composed of many cycles, each of which is executed as follows: For each output variable Y_j and for each granule belonging to the fuzzy partition that corresponds to Y_j. Iteratively, an initial fuzzy hypothesis corresponds to a combination of certain number of MFs (each of which corresponds to granule of an input variable) is created (formed) by a systematic exhaustive search procedure. Once a fuzzy hypothesis is formed it is loaded or incorporated in the hybrid fuzzy-neuro network weights for test purposes, its components (elements) will be adjusted to fit the training data. Such hypothesis is considered as a potential candidate to be a rule and then is questioned and adjusted during learning using appropriate adjustments operations by the means of a hybrid fuzzy-neuro possibilistic network using a successive approximation algorithm of systems of Min-Max relational equations. This adjustment is repeated until finding the ones that minimize the signal error. Hence another new combination is then generated and we repeat the same procedure. Thus the obtained adjusted hypotheses that minimize the cost over all possible combinations and that were embedded in the weights of the hybrid fuzzy-neuro possibilistic network are kept in a temporary learning table (for efficient storage and efficient execution).

As illustrated in Figure 5, our choice of appropriate adjustments operations which preserve interpretability and their small number only two, is not arbitrary, and indeed it is a crucial design issue, at the extreme the iterative execution of the adjustment by expanding the kernel (as in Figure 5.a., which keeps the support unchanged) converges to the upper envelope, whereas at the extreme, the iterative execution of the adjustment by shrinking the support (as in Figure 5.b., which keeps the kernel unchanged) converges to the lower envelope, which constitute the local stopping criteria in either cases. Indeed, those two operations are complementary, are better executed in mixture consecutively and together play the role and have the effect of a sliding or a moving fuzzy window.

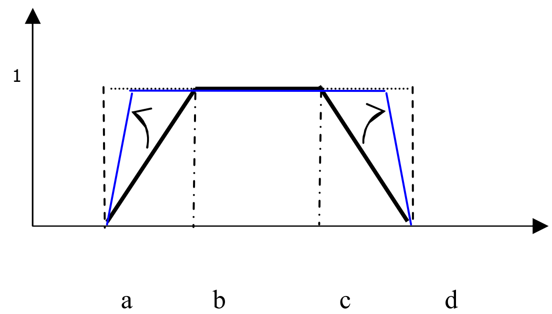

Figure 5.a: Adjustment by Expanding the Kernel

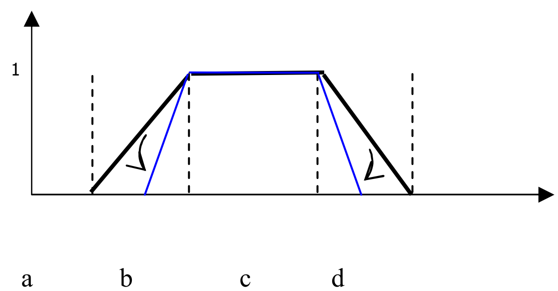

Figure 5.b: Adjustment by Shrinking the Support

The algorithm proceeds by increasing the granularity and repeats the same cycle, until reaching a satisfying point. In general the learning is stopped when either a certain level of accuracy has been reached or it is impossible or it is computationally worthless to seek minimizing the error much more, i.e. this situation means that increasing the granularity is no more interesting. In general this point constitutes a trade-offs between tractability and low cost solution. Learning need to find an approximate solution that is not necessarily precise (or crisp) optimal one but at the same time it builds a model that do manage to resolve the problem at hand effectively. At the end one or more of the obtained adjusted hypotheses that minimize the cost (over all considered granularity levels) constitutes a valid hypothesis and is transferred and stored in a knowledge base (KB) of the system as it consists effectively of a new learned rule. The system check whether or not a rule is new, i. e. whether or not it is already included the KB, and if necessary, transmits it to the KB, in an intelligible form for the storage (hash table data structure). Assume the system get two or more valid hypotheses, after checking each one, each one is eventually added to the KB as a new rule. The advantage is that by construction (learning) we build a production system with no contradictory rules and thus giving a high satisfactory performance. This is in fact a built-in quality attribute. When learning completes, we get a set of rules representing policies and heuristic strategies.

Thanks to these granular functionalities, this novel learning algorithm constitutes a departure from the conventional ones, in that it conjointly determine dynamically during the learning-time the required satisfying number of rules necessary to model the problem as well as the rules themselves explicitly. Intuitively, this number is proportional to the degree of complexity of the problem at hand.

The resolution of fuzzy relations equations constitutes a good tool in fuzzy modeling especially for dealing with inverse problems. The *fuzzy relational calculus* theory (Di Nola et al. [47], Beldjehem [13]) provides us with some theorems and a set of analytic formulas expressing solutions for some types of equations and their systems. However, usually the system is inconsistent and the existence of solutions of the system is not known in advance. This makes any preliminary analysis rather tedious if not impossible. We reformulate the problem of solving a system of Min-Max from interpolation-like format to approximation-like one. This means that instead of trying to find exact solution, we try to find the best approximate solution. Any scalar and any element of vectors or matrices are assumed to have its value in the interval [0, 1]. Formally; our problem can be stated as follows: "Given an **m x n** matrix **R** and an **n** vector **b**, find an m vector **a** such that (**a Δ R ⊇ b**) where **Δ** is the Min-Max composition and ⊇ denotes the fuzzy inclusion operation. Let us consider the case when there is no solution for the system (it does not satisfy the necessary condition, i.e. **a Δ R ⊃ b).** This can be also reflected by only computing a distance. Let A, A' be fuzzy subsets of U and α, α' be the corresponding grades of membership vectors. By ||α-α'|| we denote the number Max ($|\alpha_i - \alpha'_i|$) over i, i.e. the maximum of the absolute values of the differences between all element of α and α'. It might be interpreted as the signal error subject to be minimized. Equivalently by using this distance rather than the fuzzy inclusion concept we get the same results; and for this reason we use such a distance || **<u>a</u>** Δ R - **b** || in our implementation of the system. It corresponds to minimal distance, hence **<u>a</u>** is the best approximator. Thus, since our algorithm is valid for both interpolation-like and approximation-like formats, it allows to resolve the more general following problems: "Given an **m x n** matrix **R** and an n vector b, find <u>all</u> m vectors such that **a** Δ **R ⊇ b**". This algorithm is used as approximation procedure by the learning algorithm in our system. The learning consists mainly in crunching (approximating) systems of Min-Max equations while manipulating abstract synthetic linguistic concepts (labels, hypotheses). It can be shown that the best approximator (from the fuzzy inclusion point of view) corresponds to the lower bound **<u>a</u>** of the inf-semi-lattice. It can be computed straightforwardly using the ε resolution operator only. It has been shown by a worst-case analysis that our computing algorithm has a linear complexity of Θ (m x n) Beldjehem [13] in his thesis (required in on-line learning for time-critical applications, as well in learning large-scale, high-dimensional problems). In order to illustrate the functioning and the behavior of our *approximation algorithm* let us hand-execute it on the following example, R and b are known. The ε operator is defined in Beldjehem [13] in his thesis as follows:

$$x \;\varepsilon\; y = \begin{cases} y & \textit{if} \quad x < y \\ 0 & \textit{otherwise} \end{cases}$$

$$R = \begin{bmatrix} 0.5 & 0.6 & 0.1 & 0.3 & 0.6 \\ 0.7 & 0 & 0.8 & 0.4 & 0.7 \\ 0.8 & 0.3 & 0.5 & 0.7 & 0.6 \\ 0.4 & 0.8 & 0.6 & 0.8 & 0.7 \\ 0.4 & 0.4 & 0.7 & 1 & 0.6 \\ 0.9 & 1 & 1 & 1 & 0.8 \end{bmatrix}$$

$$b = \begin{bmatrix} 0.3 & 0.3 & 0.5 & 0.4 & 0.5 \end{bmatrix}$$

Firstly, we compute the lower bound **<u>a</u>** of the inf-semi-lattice

<u>a</u> =∨ (R ε b), where ∨ stands for MAX

<u>a</u> =[0.5 0.3 0 0 0 0]

By performing the Min-Max composition, we have

b =[0.3 <u>0.3</u> <u>0.5</u> <u>0.4</u> 0.5] (the target vector)

<u>a</u> Δ R =[0.4 <u>0.3</u> <u>0.5</u> <u>0.4</u> 0.6] (the actual output)

|| **<u>a</u>** Δ R - **b** ||=0.1, observe the surprising remarkable approximating power of **<u>a</u>**

3.2 *Abstract Computational Model of a Learning Session*

We are interested herein by establishing the computational abstract model of learning, learning implements a kind of successive approximation of Min-Max system process, and find weights of the hybrid fuzzy-neuro networks that fits “best” the data that consists of pairs I/O of the training set**.** Formally, from the computational point of view, for each output (s_k), a learning session consists to resolve or to approximate (r + 1) systems of Min-Max equations, as follows:

$$a \Delta R^{(0)} \supseteq b$$

$$a \Delta R^{(1)} \supseteq b$$

$$. \quad . \quad .$$

$$a \Delta R^{(r)} \supseteq b$$

Learning consists to prefer (validate) the configuration (the fuzzy hypothesis) of the best approximate solution (from the fuzzy inclusion point of view), i.e. which minimizes the local cost function and hence the corresponding deep structure. In other terms the learning process finds incrementally the "*best*" deep structure which corresponds to the following matrix $R^{(l)} : l \in [0,r]$ such that:

$$a \Delta R^{(j)} \supseteq a \Delta R^{(l)} \supseteq b, \forall j=0...r$$

Or equivalently,

$$\| a \Delta R^{(j)} - \mathbf{b} \| \geq \| a \Delta R^{(l)} - \mathbf{b} \|, \forall j=0...r$$

Learning tries progressively by *successive approximation* to minimize the local cost function by the generation and the approximation of a new system. Thus, this approximation algorithm constitutes the mathematical machinery of learning. Furthermore it is now clear that the ultimate aim of learning is to generate a consistent system which correspond to exact solution (or to establish a *universal interpolator*), however it seems that is not always the case in practical applications. In general the value of the local cost function may be seen as a *quality index* for a learning session or a *performance index* for the system. Learning has high speed due to its simplicity and analytic nature. The learning consists mainly in crunching (approximating) systems of Min-Max equations while manipulating abstract synthetic linguistic concepts (labels, hypotheses). Indeed the fuzzy learning process may be thought of as a new kind of algorithmic fuzzy optimization or rather an *algorithmic fuzzy approximation*. Moreover, learning is firmly grounded in formal *Fuzzy relational calculus*.

4. Validation and Verification Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners

4.1 *Validation versus Verification*

The problem of assuring correctness of Hybrid Granular Fuzzy-Neuro or Neuro-Fuzzy Learners is a good deal more critical for the utility industry than most other areas in that the ultimate consequences of failures of this Soft Computing technology, when accompanied by other technology breakdowns, can be nothing less than disastrous for, particularly, our nuclear plants--and much of the Soft Computing activity is in these facilities. Increasing attention must therefore be paid to Validation and Verification (V&V) issues.

The terms Validation and Verification suffer confusion among themselves and with related terms like “testing” and “evaluation.” Perhaps the clearest definitions are given in IEEE Std. 729-1983[48]:

Verification The process of determining whether or not the products of a given phase of software development meet all the requirements established during the previous phase.

Validation The process of evaluating software at the end of the development process to ensure compliance with software requirements.

Validation, strictly speaking, occurs only once—at the end of the last stage of the life-cycle—at which time the final system is compared to the original requirement. The objective of validation is not to insure the accuracy of each development stage but rather to insure that, however developed; the final product is and does exactly as it was intended in the beginning. Validation can be simulated in earlier stages to the extent that one can devise tests which address the question “if a real system were

developed and delivered according to the specifications developed so far, would this system satisfy the customer's initial requirements.
Given these explications Boehm's [49] succinct summary makes very good sense:
Verification "*Are we building the product right?*"
Validation "*Are we building the right product?*"

A Hybrid Granular Fuzzy-Neuro system (HGFN) is a fuzzy neural network that learns to classify data using fuzzy rules and fuzzy classifications (fuzzy sets). A Hybrid Granular Fuzzy-Neuro system has advantages over fuzzy systems and traditional neural networks: A traditional neural network is often described as being like a "black box", in the sense that once it is trained, it is very hard to see why it gives a particular response to a set of inputs. This can be a disadvantage when neural networks are used in mission-critical tasks where it is important to know why a component fails. Fuzzy systems and Hybrid Granular Fuzzy-Neuro systems do not have this disadvantage. Once a fuzzy system has been set up, it is very easy to see which rules fired and, thus, why it gave a particular answer to a set of inputs. Similarly, it is possible with such a system to see which rules have been developed by the system, and these rules can be examined by experts to ensure that they correctly address the problem.

4.2 *Validation of Hybrid Granular Min-Max Fuzzy-Neuro Relational Learners*

In general two options are possible to validate hybrid granular soft computing models, empirical validation or formal validation. One has to distinguish and contrast, between empirical validations techniques such as those based on *cross-validation* (Leave one-out, and K-folds variants) which is merely a statistical method devised to assess the accuracy of a classifier, and rigorous formal systematic ones which are established to study and assess the *generalization capacity* of a model mathematically. Exhaustive empirical validation is impossible and is meaningless when using biased Validation Set.

The second option is adopted herein to validate our Granular Min-Max Fuzzy-Neuro Relational Learners in this section of our present paper. The first one is *ad hoc* and empirical approach, that consists to use a Test Set and a Validation Set of examples in order to validate the system, the second that we have adopted herein is a rigorous systematic and formal approach which attempts the validate the system by studying *value approximation* from the perspective of generalization, as it is accepted that *generalization capability* is proportional to *value approximation*, which is in conformance with the well-known *Inductive Learning Hypothesis-ILH* in machine learning.

The notion of fuzzy set itself support by nature the idea of approximation; it is clear that in practice, when constructing fuzzy sets or fuzzy relations, the membership function approximates what is subjectively considered to be the relation of each member of a universe of discourse with a linguistic value of a variable or relation. It has been noticed that small deviations from what might be considered as "precise membership value" should normally be of no practical significance. This remains true when computing possibility/necessity measures. Informally a high-quality HFN system is a system that equates the performance of human experts. In this part we are interested to propose a validation approach to Hybrid Granular Fuzzy-Neuro Min-Max Learners for which we have already proposed and developed a fast learning algorithm that proceeds by *successive approximation* of Min-Max systems of fuzzy relational equations in the first part above. This model uses Possibility/Necessity measures and Min-Max compositional rule to perform the inference for a given input pattern. In our current study our focus is on analysis of the input-to-output mapping performed by the Hybrid Granular Min-Max Fuzzy-Neuro system. The validation bring us to study the *value approximation* of the system, the notion of approximation of the values of fuzzy variables is discussed. In our current work a *measure of proximity* of fuzzy values is introduced and a definition of fuzzy equality or *approximately equal fuzzy values* is given. We will show that the *property of approximation* is preserved when the Min-Max compositional rule is applied during inference by the fuzzy-neuro network, when used with so defined approximately equal fuzzy values, gives approximately equal results. Thus this rule preserves the property of approximation when it is applied to entities characterized by approximately equal fuzzy values.

4.2.1. *Definition*

Let A, A' be fuzzy subsets of U and α, α' be the corresponding grades of membership vectors.

By $\|\alpha-\alpha'\|$ we denote the number $\max_i(|\alpha_i-\alpha'_i|)$, i.e. the maximum of the absolute values of the differences between all element of α and α'. We define the *approximate or fuzzy equality* as follows:

A and A' are said to be *approximately equal* (and this is denoted by A ≈ A') iff given a small nonnegative number ε, it is $\|\alpha-\alpha'\| \le \varepsilon$.

The number ε is said to be a proximity *measure* of A and A'.

4.2.2. *Two Lemmas*

Lemma 1 Let k, l, m be nonnegative numbers. Then

$$|\min(k, l) - \min(l, m)| \le |k-m| \quad (4)$$

Proof

Table 2 shows all combinations of values of min (k, l) and min (l, m).

Case 1. min (k, l)=k imply k ≤l and min (l, m) =l imply l≤m. Hence k≤l≤m, and therefore |k-l| ≤ |k-m|

Case 2, 3. Obvious

Case 4. min (k, l)=l imply l ≤k and min (l, m) =m imply m≤l. Hence m≤l≤k, and therefore |k-l| ≤ |k-m|

Table 2

Case	min (k, l)	min (l, m)	|min (k, l)- min (l, m)|
1	k	l	|k-l|
2	k	m	|k-m|
3	l	l	0
4	l	m	|l-m|

Lemma 2. Let k, l, m be nonnegative numbers. Then

$$\max(k, l) - \max(l, m) \le |k-m| \quad (5)$$

Proof. (Similar to above)

Table 3 shows all combinations of values of max (k, l) and max (l, m).

Case 1. max (k, l)=k imply k ≥l and max (l, m) =l imply l≥m
Hence k≥l≥m, and therefore |k-l| ≤ |k-m|

Case 2, 3. Obvious

Case 4. max (k, l)=l imply l ≥k and max (l, m)=m imply m≥l
Hence m≥l≥k, and therefore |k-l| ≤ |k-m|

Table 3

Case	max (k, l)	max (l, m)	|max (k, l) - max (l, m)|
1	k	l	|k-l|
2	k	m	|k-m|
3	l	l	0
4	l	m	|l-m|

4.2.3. *Some Proprieties of Approximately Equal Fuzzy Values*

Let A, A' and B be fuzzy subsets of U and A ≈ A'. Then the following properties are true.

Property 1 $A \cap B \approx A' \cap B$ (6)

Proof Let $\alpha = (\alpha_i)$, $\alpha' = (\alpha'_i)$ and $b = (b_i)$ be the grades of membership vectors corresponding to A, A', B respectively. Because of Lemma 1 we have

$$|\min(\alpha_i, b_i) - \min(b_i, \alpha'_i)| \le |\alpha_i - \alpha'_i|$$

Therefore,

$$\max_i (|\min(\alpha_i, b_i) - \min(b_i, \alpha'_i)|) \le \max_i (|\alpha_i - \alpha'_i|) = \|\alpha - \alpha'\|$$

And thus **(6)** follows

Property 2 $A \cup B \approx A' \cup B.$ (7)

Proof (Similar to above)

Let $\alpha = (\alpha_i)$, $\alpha' = (\alpha'_i)$ and $b = (b_i)$ be the grades of membership vectors corresponding to A, A', B respectively. Because of Lemma 2 we have

$$|\max(\alpha_i, b_i) - \max(b_i, \alpha'_i)| \le |\alpha_i - \alpha'_i|$$

Therefore,

$$\max_i (|\max(\alpha_i, b_i) - \max(b_i, \alpha'_i)|) \le \max_i (|\alpha_i - \alpha'_i|) = \|\alpha - \alpha'\|$$

and thus **(7)** follows

4.2.4. Min-Max Compositional Rule used by our Granular Min-Max Fuzzy-Neuro Relational Learner

Let A, A' be fuzzy subsets of U and B, B' be fuzzy subsets of V, with $\alpha = (\alpha_i)$, $\alpha' = (\alpha'_i)$, $b = (b_j)$ and $b' = (b'_j)$ being the corresponding grades of membership vectors.

Let R, R' be fuzzy relations from U to V with $R = [r_{ij}]$ and $R' = [r'_{ij}]$ being the corresponding grades of membership vectors.

Theorem 1

$A \approx A'$ implies $A \,\Delta\, R \approx A' \Delta\, R$ (8)

$R \approx R'$ implies $A \,\Delta\, R \approx A \,\Delta\, R'$ (9)

Δ denotes the MinMax composition.

Proof

We apply Lemma 1 and Lemma 2 and

We use **(7)** $|\min_i \{x_i\} - \min_i \{y_i\}| \le \min_i \{|x_i - y_i|\}$

Let $A \,\Delta\, R = B$ and $A' \Delta\, R = B'$. We have:

$b_j = \min_i \{\max(\alpha_i, r_{ij})\}$ and $b'_j = \min_i \{\max(\alpha'_i, r_{ij})\}$

$| b_j - b'_j | = | \min_i \{\max(\alpha_i, r_{ij})\} - \min_i \{\max(\alpha'_i, r_{ij})\} |$

$\leq | \min_i \{ | \max(\alpha_i, r_{ij}) - \max(\alpha'_i, r_{ij}) | \} |,$

$\leq | \min_i (| \alpha_i - \alpha'_i |) | \leq | \max_i (| \alpha_i - \alpha'_i |) | = \|\alpha - \alpha'\|$

It follows that

$\max_j (| b_j - b'_j |) \leq \|\alpha - \alpha'\|$

Thus $B \approx B'$ and $A \Delta R \approx A' \Delta R$.

(9) Similar to the above

Thus the Min-Max rule preserves the property of approximation when it is applied to entities characterized by approximately equal fuzzy values. Hence, using Min-Max is an appropriate choice in hybrid fuzzy-neuro or neuro-fuzzy models, as it is accepted that *generalization capability* is proportional to *value approximation*. Thus the Granular Fuzzy-Neuro Min-Max Relational Learner might be thought of as a transparent learning device of any non-linear topology preserving mapping of inputs into an output. It has been proved formally too that Min-Max composition preserves the *value approximation* property in connections with Min-Max fuzzy-neuro network relational models as well as in rule-based fuzzy systems setting. Putting it simply, if a model approximates well it will generalize well.

4.2.4. Max-Min and α-composition

Consider the dual problem of Min-Max or the Max-Min composition. By taking some precautions, it can be shown by duality:

Theorem 2 (dual of Theorem 1)

$A \approx A'$ implies $A \circ R \approx A' \circ R$ (10)

$R \approx R'$ implies $A \circ R \approx A \circ R'$ (11)

Where $\circ$ denotes the Max-Min composition

Theorem 2 can be shown easily by duality. Thus the Max-Min rule too preserves the property of approximation when it is applied to entities characterized by approximately equal fuzzy values.

Concerning α-composition (the dual operator of ε), it can be shown using counter-example (see Appendix for more details) that:

$B \approx B'$ does not necessarily imply that $R \alpha B \approx R \alpha B'$ (12)

$R \approx R'$ does not necessarily imply that $R \alpha B \approx R' \alpha B$. (13)

This means that the α-composition does not preserve the property of approximation.

One concludes that from the perspective of generalization, it is always suitable to use standards Min, and Max instead of t-norms and t-conorms, which do not preserve the value approximation property. Our model is proved mathematically to be tolerant to small changes in input. This ensures that definitely it well generalizes to patterns that are never seen in the Training Set. Surprisingly, this same property is also hold in the Kohonen Feature map model and is it called topology preserving maps.

5. Concluding Remarks and Future Work

The conception, engineering and adoption of hybrid granular soft computing learners combining several paradigms constitutes an interesting tool in fuzzy modeling, especially useful in extracting or tuning fuzzy IF-THEN rules for complex real-world problems. This hybrid methodology has recently emerged as very promising area in soft computing fields. It consists to build hybrid granular learning models combining fuzzy logic, neural networks end evolutionary algorithms. Such models rather than conventional ones are well-suited to resolve complex real-world problems, thanks to their *learning ability*, *value approximation*, *generalization capability*, *robustness*, *transparency* and *tolerance*. The main advantage of our framework is the adoption of a hybrid granular data-driven model-free approximation learning methodology, which shortens design/development time, reduce cost and allows the construction of approximate models by learning. Moreover, learning is firmly grounded on fuzzy relational calculus, linguistic approximation and the crucial notion of importance widely used in human decision making and problem-solving.

In Soft Computing, the production of cost-effective high-quality Hybrid Granular Fuzzy-Neuro or Neuro-Fuzzy Learners is a central issue toward the success and adoption of such technology. One land mark awaited in the maturing of such technology is verification and validation (V&V). Clearly, there is little benefit in employing such a complex system unless it can be trusted to perform its intended function. In our current study our focus was on analysis of the input-to-output mapping performed by the hybrid Granular MinMax fuzzy-neuro system and we proposed a rigorous systematic formal validation theory which is firmly grounded in formal mathematics. Furthermore, the results of our present work confirm that Hybrid Granular MinMax Fuzzy-Neuro Relational models are *universal approximators*, i.e., they can approximate to arbitrarily accuracy any continuous mapping defined on a compact (closed and bounded) domain. More over, Min-Max Relational Learners as well as their dual Max-Min ones are the “right models” to build from a validation perspective, or equivalently a generalization perspective.

We propose in our future work to interpret the system as a collection of IF-THEN fuzzy rules and we expect to be able to use the good techniques from V&V of rule-based systems [50] to fuzzy rules. It might appear that a fuzzy system can be verified and validated more easily than conventional rules, because a fuzzy system uses vague terms to explain the control actions, and would therefore be easier to understand and cope with. In reality the V&V problems may be more difficult. As every thing in fuzzy logic is a matter of degree, the *consistency model* valid for expert systems does not work in the case of fuzzy rules. The concept of consistency must be refined into a notion of *degree of consistency*. Therefore, building models of the hybrid granular soft computing systems and expressing them in a formal or less formal way may be a good practice. Such models can always provide an objective for the executable system. We believe that the fields of Machine learning, Software Engineering, Knowledge Engineering and Soft Computing have to learn from each other, in order to reach the maturity level required in the field of Hybrid Systems Technology. Another land mark awaited in the maturing of V&V technology is its active adoption by the industry. Although many good theoretical techniques and methods are published in the literature, one cannot go directly from published examples to more complex examples. Nowadays, systems tend to become more and more complex and therefore the abstract models that have been used so far need to be reconsidered. The idealizing assumptions made for the development of a hybrid system must deal with the complexity of the environment in which the hybrid system runs. To develop operational, dependable, and reliable systems, developers need to work harder to define the scope of the application, the limitation of the domain, the requirements, and to verify and validate rigorously those aspects of the system. It is only in the context of practical applications that the various V&V methods will reveal their true worth.

References

[1] L. A. Zadeh, “Fuzzy sets,” Info. Control, vol. 89, pp. 338-353, 1965.

[2] L. A. Zadeh, “Biological applications of the theory of o fuzzy set and systems,” In Proc. of International Symposium on Biocybernetics of the Central Nervous System (Little, Brown and Company, Boston), pp. 199-206, 1969.
[3] L. X. Wang and J. M. Mendel, “Fuzzy basis functions, universal approximation, and orthogonal least-squares learning,” IEEE Trans. on Neural Networks, vo. 3, pp. 807-814, 1992.
[4] K. Hornik and al. “Multilayer feedforward are universal approximator.” Neural Network 2, 1989, pp. 359-366.
[5] L. A. Zadeh, “Fuzzy logic neural networks, and soft computing,” Communications of the ACM, vol. 37, pp. 77-84, 1994.
[6] L. A. Zadeh, “Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic,” Fuzzy Sets and Systems, vol. 19, pp. 111-127, 1997.
[7] L. A. Zadeh, “Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information / intelligent systems,” Soft Computing, vol. 2, pp.23-25, 1998.
[8] R. Yager and L. A. Zadeh, Fuzzy Sets, Neural Networks and Soft Computing, New York: Van Nostrand Reinhold: 1994.
[9] W. Pedrycz Ed., Granular Computing: An emerging paradigm, Springer. Series: Studies in Fuzziness and Soft Computing, vol. 70, 2001.
[10] Y. Y. Yao, “Granular computing: basic issues and possible solutions,” Proc. of the 5th Joint Conference on Information Sciences, pp. 186-189, 2000.
[11] M. Gupta, L. Jin, and N. Homma, Static and Dynamic Neural Networks: From Fundamentals to Advanced Theory, New York: John Wiley and Son Inc., 2002.
[12] T. Y. Liu, Y. Y. Yao and L. A. Zadeh, “Data Mining, Rough Sets and Granular Computing,” Heidelberg: Physica-Verlag, 2002.
[13] M. Beldjehem, “Conception, Design, Development and Validation of Hybrid Fuzzy-Neuro Systems by Approximation of Relational Min-Max Systems Algorithms: the Fennec System.” Ph.D. Thesis in Computer Science and Software Engineering, Université de la Méditerranée, Aix- Marseille II, 1993 (in French).
[14] M. Beldjehem, “Fennec, un générateur de systèmes neuro-flous.” in Proc. les Actes des Applications des Ensembles Flous, Nîmes, France, pp. 209 -218, 1993(in French).
[15] M. Beldjehem, “Le système fennec.” in Electronic BUSEFAL, vol. 55, pp. 95-104, , 1993(in French).
[16] M. Beldjehem, “The fennec system.” in Proc. ACM Symposium on Applied Computing (SAC), Track on fuzzy logic in Applications, pp. 126-130, Phoenix, AZ, 1994.
[17] M. Beldjehem, “Machine Learning based on the possibilistic-neuro hybrid approach: design and implementation.” in Electronic BUSEFAL, vol. 87, pp. 95-104, 2002.
[18] M. Beldjehem, “Learning IF-THEN Fuzzy Weighted Rules.” in Proc. International Conference of Computational Intelligence, Nicosia, North Cyprus, , CD-ROM, 2004.
[19] M. Beldjehem, “Validation of Hybrid MinMax Fuzzy–Neuro Systems.” in Proc. International conference of NAFIPS, Montreal, CD-ROM, 2006.
[20] M. Beldjehem, “Towards a Validation Theory of Hybrid MinMax Fuzzy–Neuro Systems.” in Proc. the WSEAS International Conference, Sofia, 2008.
[21] Beldjehem M. 2008. “A Validation Theory of Hybrid MinMax Fuzzy–Neuro Systems,” in Proc. the CIMSA International Conference, Istanbul.
[22] M. Beldjehem, “A Unified Granular Neuro-Fuzzy Framework for Learning Fuzzy Systems.” in Proc. the NASTEC International Conference, Mc Gill, Montreal, 2008.
[23] M. Beldjehem, “A Granular Neuro-Fuzzy Framework for Recognition of Arabic Handwriting: The GoavmRec System,” in Proc. the NASTEC International Conference, Mc Gill, Montreal, 2008.
[24] M. Beldjehem, “Granular Unified Framework for a Machine Visual System,” in Proc. the NASTEC International Conference, Mc Gill, Montreal, 2008.
[25] M. Beldjehem, “A Granular Unified Hybrid Min-Max Fuzzy-Neuro Framework for Predicting and Understanding Software Quality,” in Proceedings of NASTEC’09, Atlanta, USA, 2009.
[26] M. Beldjehem, “A Granular Unified Framework for Learning of Fuzzy Systems”, in Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), Vol.13 (4), 2009, pp. 520-528.
[27] M. Beldjehem, “A Granular Framework for Recognition of Arabic Handwriting,” in Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), vol.13 (4), pp. 512-519, 2009.
[28] M. Beldjehem, “Toward a Theory of Validation of Hybrid MinMax Fuzzy Neuro Systems,” International Journal of Computing and Mathematical Applications, Vol. 3 (1-2), pp. 1-8, 2009.
[29] M. Beldjehem, “A Granular Unified Hybrid MinMax Fuzzy-Neuro Framework for Predicting and Understanding Software Quality,” in International Journal of Software Engineering and Applications, vol. 4(4), Special Issue on *Advances in Software Engineering with Soft-Computing*, (Guest Editors Jose C. Riquelme, Roberto Ruiz, and Daniel Rodriguez), pp. 1-18, 2010.

[30] N. K. Sinha and M. Gupta, Ed., Soft Computing and Intelligent Systems: Theory and Applications, New York: Academic Press, 1999.
[31] J. R. Quinlan, C 4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.
[32] Y. F. Yuan and M.J. Shaw, "Induction of fuzzy decision trees," Fuzzy Sets and Systems, vol. 69, pp. 125-139, 1995.
[33] L. A. Zadeh, "Toward a theory of fuzzy systems," in R. E. Kalman, N. Declaris, Eds., Aspects of Network and System Theory, New York : Holt, Rinehart and Winston, pp. 209-245, 1971.
[34] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," Fuzzy sets and syst., vol.1, pp. 3-28, 1978.
[35] L. A. Zadeh, "A theory of approximate reasoning," in J. E. Hayes et al. Eds., Machine Intelligence, vol. 9 Elsevier, pp. 149-194, 1979.
[36] R. Yager, Fuzzy Set and Possibility Theory: Recent Developments, New York: Pergamon Press, 1986.
[37] D. Dubois, and H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, New York: Plenum Press, USA, 1988.
[38] W. Olaf, Possibility Theory with Application to Data Analysis, John Wiley & Son Inc., 1998.
[39] G. Miller, "The magical number seven, plus or minus two," The Psychological Review, vol. 63, pp. 81-97, 1956.
[40] L. A. Zadeh, "Fuzzy sets and information granulation," in M. Gupta, R. K. Ragade, R. R. Yager Eds., Advances in Fuzzy Set Theory and Applications, North–Holland Publishing Company, pp. 3-18, 1976.
[41] R. Bartolin, "Aide au diagnostic médical par mesures de comparaisons floues et pouvoir séparateur," Thèse d'Etat en Biologie Humaine, Marseille, 1987.
[42] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," IEEE Trans. Syst. Man Cybernet., vol. 3, pp. 28-44, 1973.
[43] M. Cayrol, H. Farreny, and H. Prade, "Fuzzy Pattern Matching." Kybernetics, vol. 11, pp. 103-116, 1982.
[44] D. Dubois, H. Prade, and C. Testemale, "Weighted fuzzy pattern matching", Fuzzy sets and systems, vol. 28, pp. 313-331, 1988.
[45] R. Yager "On the interpretation of Fuzzy If-Then Rules," Applied Intelligence, vol. 6, pp. 141-151, 1996.
[46] P. P. Bonissone, "The problem of linguistic approximation in system analysis," Ph.D. dissertation, Dept. EECS, University of California, Berkeley, 1979.
[47] A. Dinola, S. Sessa, and W. Pedrycz, Fuzzy Relation Equations and Their Applications To Knowledge Engineering, Kluwer Academic Publishers, Dordrecht, NL, 1989.
[48] IEEE Standard Glossary of Software Engineering Terminology, IEEE Std. 729-1983, February 1983.
[49] B. W. Boehm, "Verifying and validating software requirements and design specifications," IEEE Software, vol. 1, no.1, pp. 75-88, January 1984.
[50] A. H. Adrian, *Intelligent Systems for Engineers and Scientists*, CRC Press, 2001.

Appendix : counter-example

R

$$\begin{bmatrix} 1.0 & 0.9 & 0.8 \\ 0.6 & 0.9 & 0.5 \\ 0.9 & 0.3 & 0.2 \end{bmatrix}$$

$$x \,\alpha\, y = \begin{cases} 1 & \text{if } x \le y \\ y & \text{if } x > y \end{cases}$$

B

$$\begin{bmatrix} 0.6 & \underline{0.8} & 0.5 \end{bmatrix}$$

B′

$$\begin{bmatrix} 0.6 & \underline{0.9} & 0.5 \end{bmatrix}$$

$$\| B - B' \| = \max_i | B_i - B'_i | = 0.1$$

$$R \,\alpha\, B = A \text{ and } R \,\alpha\, B' = A'$$

$$A = R \,\alpha\, B = \begin{bmatrix} 1.0 & 0.9 & 0.8 \\ 0.6 & 0.9 & 0.5 \\ 0.9 & 0.3 & 0.2 \end{bmatrix} \alpha \begin{bmatrix} 0.6 & 0.8 & 0.5 \end{bmatrix}$$

$$= \begin{bmatrix} \wedge(1.0 \,\alpha\, 0.6,\ 0.9 \,\alpha\, 0.8,\ 0.8 \,\alpha\, 0.5) \\ \wedge(0.6 \,\alpha\, 0.6,\ 0.9 \,\alpha\, 0.8,\ 0.5 \,\alpha\, 0.5) \\ \wedge(0.9 \,\alpha\, 0.6,\ 0.3 \,\alpha\, 0.8,\ 0.2 \,\alpha\, 0.5) \end{bmatrix}$$

$$A = \begin{bmatrix} \wedge(0.6,\ 0.8,\ 0.5) \\ \wedge(1,\ 0.8,\ 1) \\ \wedge(0.6,\ 1,\ 1) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.8 & 0.6 \end{bmatrix}$$

$$A' = R \,\alpha\, B' = \begin{bmatrix} 1.0 & 0.9 & 0.8 \\ 0.6 & 0.9 & 0.5 \\ 0.9 & 0.3 & 0.2 \end{bmatrix} \alpha \begin{bmatrix} 0.6 & 0.9 & 0.5 \end{bmatrix}$$

$$A' = \begin{bmatrix} \wedge(0.6,\ 1,\ 0.5) \\ \wedge(1,\ 1,\ 1) \\ \wedge(0.6,\ 1,\ 1) \end{bmatrix} = \begin{bmatrix} 0.5 & 1 & 0.6 \end{bmatrix}$$

$$\| A - A' \| = \vee(0,\ 0.2,\ 0) = 0.2 \geq 0.1 = \| B - B' \|$$

R

1.0	0.9	0.8
0.6	0.9	0.5
0.9	0.3	0.2

R′

1.0	0.8	0.8
0.6	0.8	0.5
0.9	0.3	0.2

$$||R - R'|| = 0.1$$

B

0.6	0.8	0.5

$A = R \; \alpha \; B$

0.5	0.8	0.6

$A' = R' \; \alpha \; B$

0.5,	1,	0.6

$$||\, A - A' \,|| = 0.2 \geq 0.1$$