A Unified Granular Fuzzy-Neuro Min-Max Relational Learner: A Case Study

Mokhtar Beldjehem

University of Ottawa, School of Information Technology and Engineering, 800 King Edward, Ottawa, Ontario, K1N 6N5, Canada
E-mail: mbeldjeh@uottawa.ca

Abstract

This paper deals with a real world problem of medical diagnosis, to this goal, we propose to learn a compact fuzzy medical knowledge base through a cognitively-motivated granular hybrid neuro-fuzzy or fuzzy-neuro possibilistic model appropriately crafted as a means to automatically extract fuzzy weighted production rules. The main idea is to start learning from coarse fuzzy partitions of the involved proteins variations of input variables and proceed progressively toward fine-grained partitions until finding the appropriate partitions that fit the data. We provide details of implementation issues, experimental results, and discussion of interpretability issues. Moreover, learning is firmly grounded on fuzzy relational calculus, linguistic approximation and the crucial notion of importance widely used in human decision making and clinical problem-solving.

Keywords: *Medical diagnosis, importance, possibility theory, if-then fuzzy weighted rules, hybrid granular fuzzy-neuro model, approximation of Min-Max relational equations, linguistic approximation.*

1. Introduction and Related Work

As early as the mid 1950s, physicians and computer scientists have acknowledged the utility of computer-aided medical diagnosis. Since then, a variety of approaches were explored and attempted, ranging from *clinical algorithms* or *flowcharts*, *pattern recognition* techniques, *decision theory*, *probabilistic* and *Bayesian* approaches to *expert-systems* or knowledge-based approaches. Each of these approaches can be applied successfully to narrow and carefully chosen medical domain. However, they suffer from serious drawbacks when applied to a broad domain of medical diagnosis. 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, and the usage of fuzzy sets theory in medical applications 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. Fuzzy-rule based systems are universal approximators Wang and Mendel [3] as multilayer feedforward neural networks 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 medical knowledge and complexities of medical diagnosis. Fuzzy rules attempt to capture the "*rules-of-thumb*" approach generally used by physicians, clinicians and biologists for clinical 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 *fuzzy-neuro* and/or *neuro-fuzzy* models as in Beldjehem [13-29], and Sinha and Gupta[30] in particular could effectively contribute to the building of next-generation intelligent reliable medical diagnosis models. The benefits from adopting such hybridization is its capacity to integrate seamlessly and to account conjointly for both empirical input-output historical data on patients (available measurement data from laboratory test results, symptoms, signs, syndromes or diseases) and heuristic domain knowledge that is available from biologists, clinicians, physicians, good practices and body of knowledge of the medicine field 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 medical diagnosis model by supervised learning from I/O examples (representing instances of the system's behavior). Besides integrating non-linearities directly from the learning examples (training set), the additional advantages of such an approach is the inherited property of *value approximation* and *robustness* which is of paramount importance in exhibiting generalizations necessary to process unseen situations (including testing set and validation set).

On one hand, this has led to intensive research for the development of accurate learning models using various paradigms ranging from Statistical Bayesian approach, neural networks, clustering algorithms, decision trees by conventional ID3 and C4.5 inductive approach of Quinlan [31] to inductive 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 Medicine, (2) the incapacity to account conjointly for both historical data and available explicit and heuristic linguistic medical 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 physicians, biologists and clinicians in their practice in regards to clinical decision making for the purposes of medical diagnosis. On the other hand due to the Black-box nature of naïve models in the context of medical diagnosis, they still are built on idealizing assumptions such the independence of attributes and due to the lack of probability distribution of possible diagnoses 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.

2. A Novel Learning Methodology

2.1. Motivations for Our Learning Methodology

In general, diseases can be organized along many dimensions: They are caused by inciting cause (etiology), they act on an organ (anatomy), the body mounts a physiologic response (pathophysiology) which results in varying degrees (severity) of dysfunctions (sign and symptoms) expressed over a period of time, and s.o. We want to conduct herein a case study in connection with a clinical medical diagnosis real world problem on

Proteins/Biological Inflammatory Syndromes (P.B.I.S) that the physicians, biologists and clinicians alike are faced with in their practice. In making diagnosis, experts usually rely on their experiences and intuition rather than applying precise step-by-step algorithms. It is well known that medical diagnosis 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 assign the appropriate diagnoses (diseases or syndromes) given patient's measurement data or patient profile (laboratory test results) about a certain number of symptoms under scrutiny. To this end, we propose to adapt and use a granular hybrid Min-Max fuzzy-neuro relational model for automatic building and assessing of such a diagnosis system. More specifically, we will use our developed hybrid fuzzy-neuro model in order to evaluate the impact of symptoms aspects on the diagnoses of a given patient, and to study the relevance of using those symptoms as indicators to characterize diagnoses (syndromes). It is worth mentioning that the concept of protein variation itself is a matter of degree and indeed could be adequately captured as a fuzzy concept. More specifically we will investigate the possibility to learn and understand causal relationships between some symptoms, and the syndromes of patients. 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 proteins in relation with the diagnosis of the patient. 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 diagnosis, and how much it is necessary that that we infer a given diagnosis. We believe that our approach will definitely open the door for next generation intelligent medical diagnosis systems. Besides machine learning the causal relationships between the symptoms and the diagnoses, they allow the detection of the importance or relevance of each symptom to diagnosis which is of paramount importance for an empirical approach of studying and understanding evidence-based medical diagnosis aspects and hence providing justification facilities for the validation issues and during the deployment and operation in the field. Thus enables the understanding of relative importance of each symptom and its influence in causing the appearance of diseases or diagnoses.

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.

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, formally validated and deployed a hybrid fuzzy-neuro system called Fennec Beldjehem [13-29] that was successfully applied to image processing and vision engineering [24] as well as to a complex handwriting pattern recognition problem [27]. Based on our previous work in connection with granular soft computing and software quality prediction [26, 29]. We first review our model, and we describe the Proteins/Biological Inflammatory Syndromes (P.B.I.S) 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 medical diagnosis by integrating some useful concepts from the human cognitive processes and adding some interesting granular functionalities and knowledge of the medical domain. In general, Medicine activities are knowledge intensive and medical diagnosis is a good application area since the knowledge available is generally heuristic in nature and in making decision clinicians, biologists and physicians 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 satisfying solution and so stops and does not need any more details.

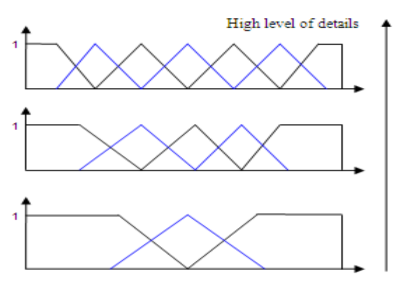


Figure 1. From a Coarse Fuzzy Partition to a Fine-Grained Fuzzy Partition

At this point, an excess of precision is not needed (is not necessary) because a certain satisfying trade-offs between 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 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.

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].

3. The Statement of the Learning Problem

3.1. Modeling of the Medical Diagnosis Problem

This framework will be described and illustrated in the medical domain with a medical application (medical diagnosis assistance from inflammatory-syndromes/proteins profiles). Conventionally, physicians, clinicians and biologists define ranges or numerical intervals in expressing proteins variation. They define and determine first normal or non pathological states, and then use them as references to which abnormalities are specified. Boundaries are in fact inaccurate to some extent as they are chosen arbitrarily and in practice are context-dependent, situation-dependant and physician-dependent. As a result, any system built on those assumptions will be faced by the boundary problem due to the usage of thresholds in defining those intervals. To remedy such a problem, it seems appealing to use and adopt Fuzzy set theory, and more specifically linguistic variables which are interpreted by labels of fuzzy sets (as shown in figure 3 for the protein Haptoglobin in relation with Vascultis). The proteins /B.I.S problem can be stated as follows: Given a protein profile composed of five normalized values (measurement on collected data of a patient represented by the input pattern to the network). Assign the appropriate diagnosis groups (or B.I.S) from eleven groups (represented by the output of the network). The fact that the groups are not mutually exclusive, add more complexity to this problem, which is obviously not a classification problem (at least from the sense of classical crisp classification perspective).

The biological data of our case study of this application of inflammatory protein variation are reported from Bartolin [41], the following five proteins variations (captured by five input neurons in our model), involved in biological inflammatory reactions, have been chosen.

- C3 (C3-Complement Fraction)
- A1AT (Alpha-1-Antitrypsine)

- Om (Orosomucoid)
- Hpt (Haptoglobin)
- CRP (C-Reactive Protein).

The Protein Biological Inflammatory Syndrome (P.B.I.S.) pattern contains eleven groups (captured by eleven output neurons in our model):

- Normal condition
- Eight Biological Inflammatory Syndromes :
    - Bacterrial infections
    - Viral infections
    - Vasculitus
    - Nephrotic syndromes
    - Acute Glomerular Nephritis
    - Intravascular Hemolysis with inflammation
    - Collagen Disease non Lupus and without infection
    - Lupus
- Intravascular Hemolysis withount inflammation
- Glomerular Renal Insufficiency without inflammation

The complexity of proteins/ B.I.S problem is due to the complexity of the inflammatory process itself from a biological perspective. The main idea in learning is to partition the input space into fuzzy regions taking into account conjointly both the generated fuzzy judgment (explicit or heuristic knowledge) and the training set (empirical knowledge). More specifically we will investigate the possibility to learn causal relationships between some symptoms, and the diagnoses (syndromes or groups). 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 proteins in relation with the diagnosis. 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). We believe that our approach will definitely open the door for intelligent next generation Medical diagnosis systems. Besides learning the causal relationships between the diagnoses and the symptoms, they allow the detection of the importance and/or relevance of each symptom in relation to diagnoses which is of paramount importance for an empirical approach of studying for understanding Biological aspects and hence providing justification facilities for the symptoms validation issues. Thus enables the understanding of *relative importance* of each symptoms and its influence in inferring diagnoses. Ultimately, this enables us to determine the *minimal subset of most relevant symptoms* allowing predicting the possible diagnosis accurately. The inputs neurons are herein a certain number of symptoms (corresponding to five proteins in our case study), our aims is at assigning the appropriate diagnoses to a patient among a certain number of possible diagnoses (eleven possible diagnoses in our case studies).

Eleven output neurons are required in order to represent the eleven possible diagnoses.

We assume herein for convenience that an inflammatory protein variation is a linguistic variable that might have linguistic values expressing and characterizing normality and variations for abnormalities 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, "Haptoglobin 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 protein variations are associated and interpreted by a *fuzzy partition* or equivalently a fuzzy sequence comprising a certain number of terms or granules depending on the protein under scrutiny, the number of which as well as the slopes of membership functions are will be determined one the a learning session is completed.

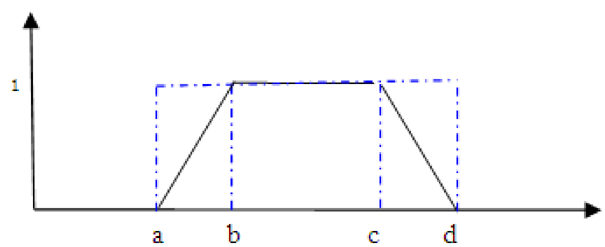


Figure 2. A trapezoidal MF as being modeled in terms of two envelopes, lower and upper envelope approximators

3.2. Description of the Learning Process

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:

- 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 (n=5 input neurons in our model for our current case study).

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.

4. Formulation of the Learning Problem

4.1. Hypothesis Generation, Formulation and Testing

During learning-time, one and only one operator which is *Interpretability-Preserving* 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 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. 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, which ensures completeness.

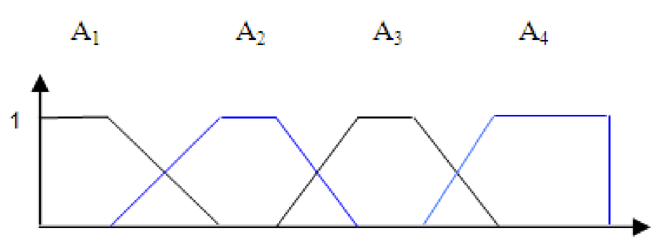


Figure 3. A Fuzzy Partition of Granularity c=4, that is a Superposition of Two Wave Functions Representing Inflammatory Protein Variation.

When the possible values for a variable are symbolic rather than numeric, which is the case of a protein variation in our medical diagnosis, linguistic approximations can be learned and represented in terms of labels of fuzzy sets with a corresponding membership functions (MFs). 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 on order to insure interpretability. $A_1, A_2 \ldots A_i \ldots A_c$ 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:

A trapezoidal MF is expressed explicitly as follows:

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

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

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

4.2. Learning by Hybrid Min-Max Fuzzy-Neuro Network

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 a fuzzy-neural possibilistic network 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.

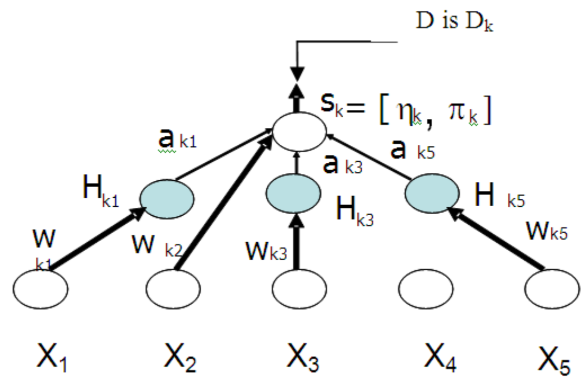


Figure 1. Schematic Representation of the Hybrid Fuzzy-Neuro Possibilistic Min-Max Model Used

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 (Primary Linguistic Weights): 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 (Secondary Interval Weights): 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 $[w_{kj} \cap \text{Not } X_j]$ be necessity measure associated to fuzzy sets w_{kj} and X_j . In general our model is governed by the three 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.

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

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

$$s_k = [\eta_k, \pi_k] \quad (3)$$

Obviously, each output variable will be assigned an interval as illustrated in equation 3; the inputs of the fuzzy-neuro networks represent the five patient's measurement data of involved proteins and eleven output variables are required in order to represent diagnoses groups 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).

Observe that Maximum ($\vee$) limits lower amplitudes of inputs, we have ($\prod(X_j; wkj) \vee a_{kj}$) $= a_{kj}$ if $\prod(X_j; wkj) \leq a_{kj}$, and amplifies higher ones ($\prod(X_j; wkj) \vee a_{kj}) = \prod(X_j; wkj)$, if $\prod(X_j; wkj) \geq a_{kj}$, 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 map is *topology-preserving* and 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 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. See TABLE IV, which reflects the symptoms' influence in relation with the presence of a given diagnosis (disease) in our framework. In connection with our problem of medical diagnosis, semantically, missing edge reflects the non-influence of the input (of the corresponding symptom) in the appearance or presence of the output (the diagnosis or disease) of the class. This enables us to determine the minimal subset of the most relevant symptoms allowing to infer the diagnosis or to identify the disease, in a fast, transparent, accurate and faithful fashion, thus enabling the understanding of the medical diagnosis intricacies and the clinical problem-solving complicated issues.

Beyond medical 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 medical 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

Thus the fuzzy-neuro possibilistic network 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 Beldjehem [19-21, 28] in connections with Min-Max fuzzy-neuro network relational models as well as in rule-based fuzzy systems setting.

5. Resolution of the Learning Problem

5.1. 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. As illustrated in Figure 4.a and Figure 4.b, our choice of appropriate adjustments operations which preserve interpretability and their small number only two, their ease of implementation, 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 4.a., which keeps the support unchanged) converges to the upper envelope (which is a crisp set), whereas at the extreme, the iterative execution of the adjustment by shrinking the support (as in Figure 4.b., which keeps the kernel unchanged) converges to the lower envelope (which is a crisp set), 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.

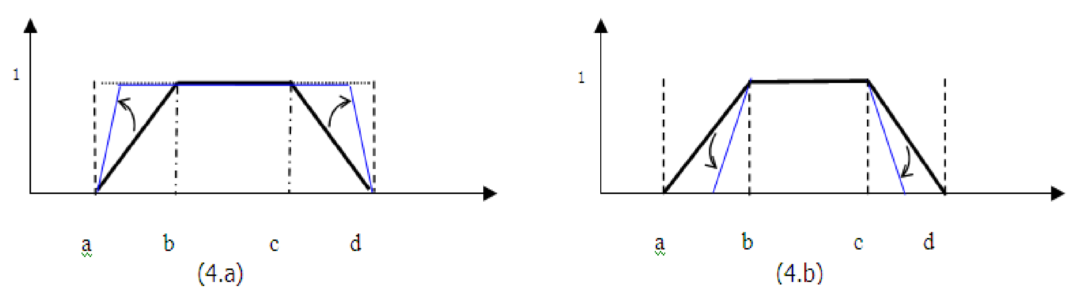


Figure 4. The two adjustment operators used (4.a): Interpretability-preserving adjustment by expanding the kernel (4.b): Interpretability-preserving adjustment by shrinking the support

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.

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 a set of analytic formulas expressing solutions for some types of equations and their systems. However, 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 ($\mathbf{a} \Delta \mathbf{R} \supseteq \mathbf{b}$) where Δ is the Min-Max composition and $\supseteq$ 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. $\mathbf{a} \Delta \mathbf{R} \supset \mathbf{b}$**).** 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 $\|\alpha-\alpha'\|$ 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 $\| \underline{\mathbf{a}} \Delta R - \mathbf{b} \|$ in our implementation of the system. It corresponds to minimal distance, hence $\underline{\mathbf{a}}$ 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 $\mathbf{a} \Delta \mathbf{R} \supseteq \mathbf{b}$". 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 $\underline{\mathbf{a}}$ 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. 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.

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

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

The ε operator is defined in Beldjehem [13] in his thesis as follows:

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

Firstly, we compute the lower bound **a** of the inf-semi-lattice.

a $= \vee$ (R ε b), where $\vee$ stands for MAX

a =[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)

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

|| **a** Δ R - **b** ||=0.1

Observe the surprising remarkable approximating power of **a**

5.2. 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:

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

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

$$\cdot \quad \cdot \quad \cdot$$

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

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. It has been shown that this system is a *universal approximator* (Beldjehem 2006, 2008). 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*.

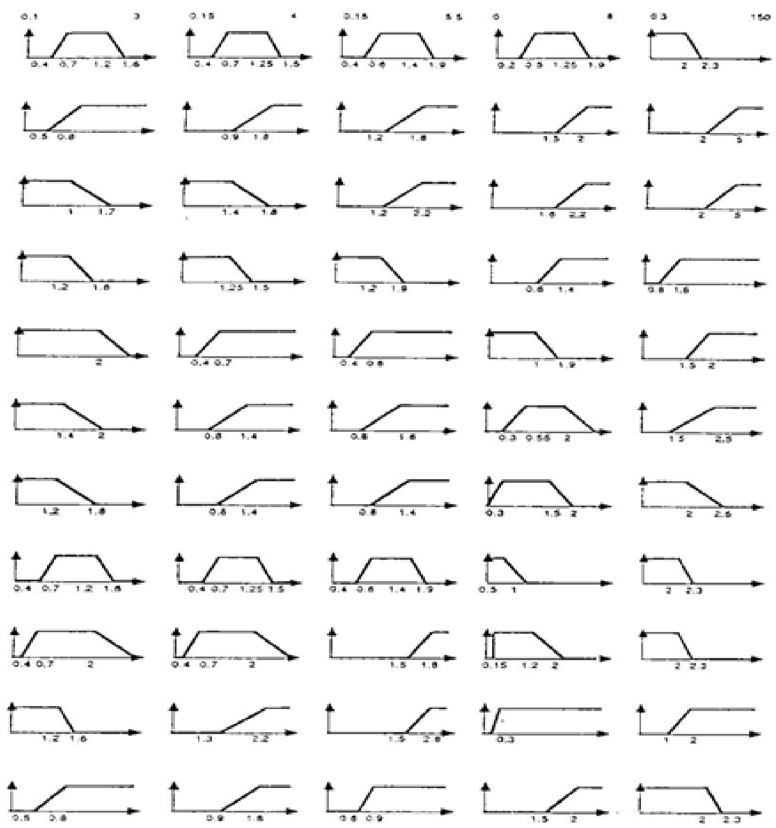
6. Experimental Results, Discussion and Interpretability Issues

We have been interested with an application to real-world data, in order to study the general behavior of the system and to illustrate its possibilities. The complexity of proteins/ B.I.S problem is due to the complexity of inflammatory process. The proteins /B.I.S problem can be stated as follows: Given a protein profile composed of five normalized values (represented by the input pattern to the fuzzy-neuro network). Assign the appropriate diagnoses groups (or B.I.S) from eleven groups (represented by the outputs of the fuzzy-neuro network).

Table 2: The Learning Table Reflecting Secondary Interval Weights

	H_{i1}	H_{i2}	H_{i3}	H_{i4}	H_{i5}
S_1	[0.5,0.5]	[0.0,0.0]	[0.0,0.0]	[0.0,0.0]	[0.0,0.0]
S_2	[0.8,0.8]	[0.4,0.4]	[0.0,0.0]	[0.0,0.0]	[0.4,1.0]
S_3	[0.9,0.9]	[0.0,0.0]	[0.2,0.2]	[0.7,0.7]	[0.2,0.2]
S_4	[0.8,0.8]	[0.2,0.2]	[0.0,0.0]	[0.2,0.2]	[0.8,0.8]
S_5	[0.8,1.0]	[0.3,0.3]	[0.3,0.3]	[0.0,0.0]	[0.2,0.2]
S_6	[0.3,0.3]	[0.3,0.3]	[0.2,0.2]	[0.8,0.8]	[0.0,0.0]
S_7	[0.3,0.3]	[0.2,0.2]	[0.2,0.2]	[0.7,0.7]	[0.0,0.0]
S_8	[0.3,0.3]	[0.3,0.3]	[0.3,0.3]	[0.0,0.0]	[0.0,0.0]
S_9	[0.3,0.3]	[0.3,0.5]	[0.1,0.1]	[0.3,0.3]	[0.0,0.0]
S_{10}	[0.1,0.1]	[0.0,0.0]	[0.0,0.0]	[0.0,1.0]	[0.2,0.5]
S_{11}	[0.5,0.5]	[0.0,0.0]	[0.0,1.0]	[0.0,0.0]	[0.0,0.0]

Table 3: The Learning Table Reflecting MFs of Primary Linguistics Weights



The training set is composed of 163 I/O examples extracted directly from historical patient records. We give herein the learning tables obtained (Table 3 and Table 4) at the end of learning session by presenting 163 I/O examples (tuples) of the problem to the fuzzy-neuro network. Simulations show that by increasing training set cardinality, extracted fuzzy rules became progressively discriminated and finally stabilize (to each output cell one and only one fuzzy rule is associated, and most numerical weights stabilize with numerical scalar values rather than an interval as shown in Table 2). This confirms our intuition that the fuzzy-neuro network tries to satisfy constraints imposed by examples. In general the value of entropy decreases progressively when increasing the number of examples. Learned membership functions (MFs) associated to linguistic primary weights are illustrated in Table 3.

The main idea in learning is to partition the input space into fuzzy regions taking into account conjointly both the generated fuzzy judgment (explicit heuristic knowledge) and the training set (empirical knowledge), this is the main advantage of the hybrid granular fuzzy-neuro modeling approach. Once learning is completed, it is possible using a *linguistic approximation* to build automatically completely a true linguistic fuzzy system by learning.

In the case under study, the fuzzy-neuro network learns exactly eleven fuzzy weighted rules, and to every possible diagnosis exactly one and only one rule is associated. Obviously, it learned a compact set of rules.

Linguistic approximation in connection with fuzzy modeling was first studied and applied by Bonissone [46] in his thesis. This task starts by identifying the membership function associated to NORMAL for each protein variation in collaboration with physicians, and then use them as references to which abnormalities are specified by generating and assigning appropriate labels to the rest of unlabeled known membership functions associated with the corresponding secondary weights of the fuzzy-neuro network (or attributes). For instance, given the NORMAL granule, by linguistic approximation, one could generate and obtain the actual variation of Haptoglobin is VERY INCREASED in relation with Vascultis from the membership function as illustrated in Figure 5.

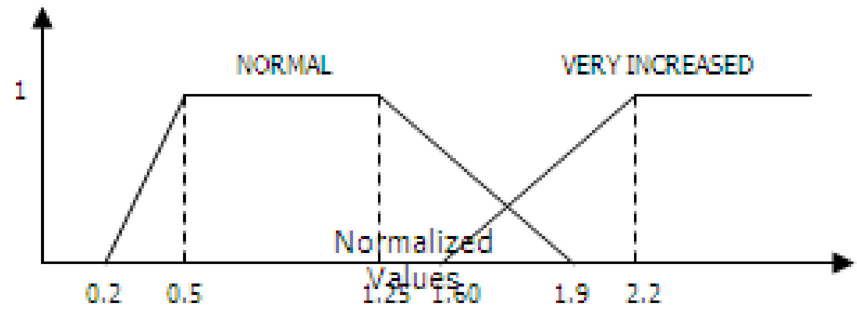


Figure 5: Linguistic Approximation of Actual Variation of Haptoglobin (X4) is in Relation with Vascultis (S3)

Intervals reflecting non-importance are approximated linguistically by means of labels as illustrated in Table 4, for instance, from Table 2, we read A_{43}= [0.7,0.7], then from Table 4, it is close to [0.6, 0.8[, thus By learning one infers that Haptoglobin (X_4) is VERY INCREASED is unimportant in relation with Vascultis (S_3), it is now clear that basically learning tries to capture this notion of importance of proteins variation (or symptoms) in relation to group diagnoses (syndromes or diseases).. Care thus needs to be taken in ensuring that biologists, clinicians and physicians understand and are satisfied with this notion of importance used in the model. We write it, (Haptoglobin (X_4) is VERY INCREASED, 0.7), in relation with Vascultis (S_3).

The effect of introducing the relative importance on the associated membership function is illustrated in Figure 6, which is similar to the application of a fuzzy cut (not a conventional crisp cut!), due to the application of the Maximum operator.

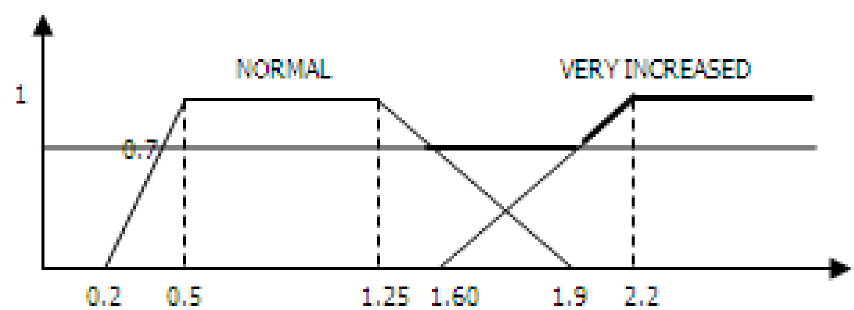


Figure 6. The Effect of Introducing the Relative Importance of Variation of Haptoglobin (X4) is in Relation with Vascultis (S3)

Table 4: The Symptom's Relative Importance in Relation with a Given Diagnosis (disease)

The magnitude of the symptom's importance	Non-importance
Not at all important	[0.9, 1]
Definitely unimportant	[0.8, 0.9[
Unimportant	[0.6, 0.8[
Neither important nor unimportant	[0.4, 0.6[
Important	[0.1, 0.4[
Very important	[0, 0.1[
Don't know (partial ignorance)	[0, 1]

8. Concluding Remarks and Future Work

We have accommodated our cognitively motivated granular computational framework for learning fuzzy systems and have illustrated how to use and apply it properly and effectively in order to resolve a complex medical diagnosis real world problem. This allows the automatic learning of fuzzy if-then diagnosis rules of systems which are large scale, too complex or too ill-defined to admit of precise quantitative analysis, description or quality control strategy. It may be thought of as an automatic means or a learning device for capturing the description of ill-defined concepts, relations and decisions rules. Such a framework integrates conjointly both the perceptual and the cognitive aspects of the clinical problem-solving process and ensure a granular processing of the underlying input from different granularity levels. The "*good*" prediction rule-base (RB) is obtained automatically from I/O training examples. Its inference engine has the inherent ability to generalize, which permit it to classify unseen examples accurately. During learning-time the system finds automatically the adequate levels of details (granularities) for the problem at hand. The main advantage of our framework is the adoption of a hybrid granular data-driven model-free approximation methodology, which shorten design/development time 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 clinical problem-solving.

Another promising alternative that constitutes a candidate solution and that we are currently exploring is to use *evolutionary algorithms* (EAs) as in (Karr. and Gentry [48], Pedrycz [49], Falkenauer [50], and Cordon et al. [51]; EAs are optimization techniques based on the mechanics of natural selection and natural genetics. They were first described by John Holland in the 1970s [52]. EAs has a great power for global optimization and do not need to know the model previously. EAs also do not require the continuity of the parameters. Therefore EAs can easily handle the multi-parameter problems of medical diagnosis and for this reason it seems appealing and convenient to use EAs too in our framework. Thus an EA may replace the generator of hypotheses subsystem in our framework. Instead of using an exhaustive search to generate all possible fuzzy hypotheses to test, it may be possible to use an EA that converges to the "*best*" global hypothesis by evolution rather than trying all possibilities. EAs can effectively contribute significantly in our framework thanks to their learning and optimization capabilities. Of particular interest to us is to devise an *evolutionary algorithm-based fuzzy system* and we are exploring its architectural, algorithmic, conception aspects and design-tradeoffs such as to try to fuzzify concepts used by EAs to obtain and use *fuzzy fitness* functions (or *fuzzy cost*), *fuzzy crossover*, *fuzzy mutation* and so on to ensure *smooth evolvability* during learning, evolution and maintenance.

Our methodology will definitely open the door for next generation intelligent medical diagnosis systems. Besides learning the causal relationships between the symptoms and the diagnoses, they allow the detection of the importance and/or relevance of each symptom to diagnoses which is of paramount importance for an empirical approach of studying for understanding medical diagnosis and hence providing justification facilities for the *validation* issues as well as for the *deployment* and operation phase. Thus enables the understanding of relative importance of each symptom and its influence in relation to diagnoses. Ultimately, this enables us to realize dimension reduction and determine the *minimal subset* of the most relevant symptoms allowing characterizing diagnoses in a fast, accurate and faithful fashion.

In general, for difficult clinical problems one could not assume that the list of disease is exhaustive and mutually exclusive. Furthermore, one could not take for granted the assumption of conditional independence of symptoms (inputs).

The simulation of human expertise is, however, not the primary goal of the field. Rather, the primary goal of this field is to develop software that performs efficiently and competently, and is able to interact and explain its reasoning and conclusions to its users (physician) in a natural way. In other terms, in most clinical situations the goal is to assist the clinicians in resolving and understanding difficult clinical problems, not to replace the clinicians in their practice.

The construction of cognitively-motivated approximate models by the integration of hybrid granular soft computing and common sense knowledge as advocated by Zadeh[7, 53] with the principles of cognitive systems based on adaptive algorithms of Holland and Reitman [54] and/or theories of cognition of Newell and Simon[55], and Newell[56] in general, evolutionary algorithm-based fuzzy systems and granular Min-Max fuzzy-neuro relational systems in particular are good candidate in order to build next generation of intelligent medical diagnosis systems. Furthermore, by mimicking the way clinicians perform diagnosis and human information processing characteristics, they are better able to communicate their reasoning to their users. Such systems are capable of learning and reasoning under uncertainty, and exhibit performance-accuracy trade-offs, adaptability, transparency, interpretability, robustness, tractability, tolerance for uncertainty, categorization abilities, value approximation and therefore ensuring smooth evolvability and generalization capacities which are required in coping with hard and difficult clinical problems of medical diagnosis. These and other similar data-driven model-free approximation approaches give promise of being more acceptable as good candidates to create programs that perform like experts and thus playing a larger role in the everyday practice of medicine.

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, Development, Validation and Evolution of Hybrid MIN-MAX Fuzzy-Neuro Systems by Approximation Algorithms of MIN-MAX Relational Systems: The Fennec System." Ph.D. Thesis in Computer Science and Software Engineering, University of Mediterranean (Aix- Marseille II), 1994 (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] C. L. Karr and E. J. Gentry, "Fuzzy control of pH using genetic algorithms." IEEE Trans. On Fuzzy Systems, vol. 1pp. 46-53, 1993.
[49] W. Pedrycz, Ed., Fuzzy Evolutionary Computation, Kluwer Academic Publishers, 1997.
[50] E. Falkenauer. Genetic Algorithms and Grouping Problems, John Willy and Sons, 1998.
[51] O. Cordon, F. Herrera, F. Hoffman, and L. Magdalena, Genetic Fuzzy System, Evolutionary Tuning and Learning of Fuzzy Knowledge Bases, World Scientific, 2001.
[52] J. H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.
[53] L. A. Zadeh, "A theory of commonsense knowledge." In H.J. Skala, S. Termini and E. Trillas; Eds., Aspect of Vagueness, Dodrecht: Reidel, pp. 257-295, 1984.
[54] J. H. Holland, and J. S. Reitman, "Cognitive systems on adaptive algorithms," in D. A. and Hays-Roth, F. Eds., Pattern-directed inference systems, waterman, Academic Press, pp. 313-329, , 1978.
[55] A. Newell and H. Simon, Human Problem Solving, Englewood Cliffs, NJ: Prentice Hall, 1972.
[56] A. Newell Unified Theories of Cognition, Harvard University Press, 1990.