Published in the *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 4, pp. 422-440, 1999.
This paper can be downloaded from ARL's Web page: http://skiron.control.ee.auth.gr/

Learning in the Framework of Fuzzy Lattices [1]

Vassilios Petridis and Vassilis G. Kaburlasos

Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki
GR-54006 Thessaloniki, Greece

Abstract

A basis for rigorous versatile learning is introduced theoretically, that is the *framework of fuzzy lattices* or *FL-framework* for short, which proposes a synergetic combination of fuzzy set theory and lattice theory. A fuzzy lattice emanates from a conventional mathematical lattice by fuzzifying the inclusion order relation. Learning in the FL-framework can be effected by handling families of intervals, where an interval is treated as a single entity/block the way explained herein. Illustrations are provided in a lattice defined on the unit-hypercube where a lattice interval corresponds to a conventional hyperbox. A specific scheme for learning by clustering is presented, namely s- *fuzzy lattice learning scheme* or s-*FLL* (*scheme*) for short, inspired from the adaptive resonance theory (ART). Learning by the s-*FLL* is driven by an *inclusion measure* σ of the corresponding Cartesian product to be introduced herein. We delineate a comparison of the s-*FLL* scheme with various neural-fuzzy and other models. Applications are shown to one medical data set and two benchmark data sets, where s-*FLL*'s capacity for treating efficiently real numbers as well as lattice-ordered symbols separately or jointly is demonstrated. Due to its efficiency and wide scope of applicability the s-*FLL* scheme emerges as a promising learning scheme.

Index Terms **-** Lattice theory, fuzzy set theory, learning systems, clustering methods.

1 INTRODUCTION

Learning has been treated in different domains such as the Euclidean space, the Boolean space, etc. As a rule, learning has been dealt with in the literature in a single domain but not in a combination of them. The present authors maintain that the capacity to consider jointly several domains, including disparate ones, could improve the efficiency of learning. We approached learning in a unifying manner by considering partly ordered sets, in particular mathematical lattices, as the learning domain. Note that a lattice does not only imply a wide scope for applicability but also it is equipped with the mathematical tools necessary for defining rigorously useful properties.

Lattices have drawn attention lately in certain scientific disciplines including logic, discrete mathematics, and computer science. Some deeper theoretical results have been considered such as the Knaster-Tarski Fixpoint Theorem [10]. Regarding applications, lattices have been employed for describing acquisition of mental models [32], and also they have been employed in fuzzy logic [12], [15]. To the best of the present authors' knowledge, the diversity of a learning domain which employs jointly several disparate lattices has not been considered in applications to-date. An older approach which deals with a lattice learning domain, specifically in neurocomputing, is noted in [22]. Nevertheless the work in [22] is primarily oriented towards medical applications rather than towards a theoretical substantiation. On the other hand the work herein maintains an established lattice theory terminology, it reaches new theoretical results, and it gains new insights while demonstrating pilot experimental results.

---

[1] This work was supported in part by European Brite Euram Project BE7470.

This work begins with the introduction of the novel notion *fuzzy lattice*, that is a lattice which fuzzifies its conventionally binary inclusion relation. A fuzzy lattice is different from an *L-fuzzy set* [16], the latter is also called *type 2 fuzzy set* [25]. An L-fuzzy set is a mapping from a universe of discourse onto a lattice and as such it is a generalization of a conventional fuzzy set which is, more strictly, a mapping from a universe of discourse onto the unit interval of real numbers [0,1]. On the other hand a fuzzy lattice stems from a "conventional lattice universe of discourse" by fuzzifying, in a conventional fuzzy set sense as it will be explained herein, its crisp binary inclusion relation. Hence the *framework of fuzzy lattices* or *FL-framework* for short, to be defined in this work, emerges as a new, hierarchic, and versatile domain for the development of learning schemes.

We have treated learning in the FL-framework herein as a procedure which defines crisp sets of points by intervals of lattice elements. However we have allowed for generalization regarding points outside crisp intervals. Note that an interval in the *N*-dimensional Euclidean space/lattice is a hyperbox. Learning schemes that define and/or enhance hyperboxes in the Euclidean space have been proposed and studied to-date in various contexts. For instance, the fuzzy adaptive resonance theory (fuzzy ART) with complement coding [9], the fuzzy ARTMAP [6], as well as the min-max neural networks [39], [40], learn either in an unsupervised or in a supervised fashion by defining and/or by enhancing hyperboxes in Euclidean space. In [2] the class of axis-parallel rectangles in *N*-dimensional Euclidean space is treated and it is shown that this class is efficiently probably approximately correct (PAC) learnable. Outside the Euclidean space, learning intervals is considered implicitly in [45] where PAC learning occurs by defining intervals in a Boolean lattice, in particular conjunctive normal forms (CNF) are learned. All the previous learning schemes consider explicitly a single learning domain, be it the Euclidean space or the Boolean space. On the other hand a learning scheme of the FL-framework is applicable in many domains, even jointly, including the previously mentioned spaces.

Several learning schemes in the FL-framework have been proposed already by the present authors. For instance in [36] the fuzzy lattice neurocomputing (FLN) scheme is presented and it is applied on benchmark data sets in the Euclidean space, in [23] the operation of the FLN is substantiated theoretically and FLN's applicability is shown in a lattice of fuzzy sets, in [37] the two-level fuzzy lattice (2L-FL) scheme for supervised clustering is presented. In the current work a lattice-applicable learning scheme is shown, that is the σ- fuzzy lattice learning (*s-FLL*) scheme, inspired from the biologically motivated Adaptive Resonance Theory (ART) [3]. ART is a well-known neurocomputing paradigm. Various models of the ART family have been presented to-date including clustering, binary [3], analog [4], [5], and fuzzy [9] versions. Supervisory learning of either binary [7] or analog [6], [8], [43], [44] Euclidean patterns has also been considered. Lately, bounds on the learning complexity of the fuzzy ART have been specified [19].

The emphasis of this paper is mainly on the theoretical introduction and substantiation of the FL-framework but we also introduce the *s-FLL* scheme. Section 2 carries the main theoretical contribution and it introduces rigorously novel concepts while it includes brief discussions of elementary lattice theory; the proofs of the new theorems and propositions are given in the Appendix. In section 3 a complete lattice in the unit-hypercube is shown and the lattice in question is decomposed into its constituent (lattice) chains, to be defined herein, in order to facilitate the analysis. We decided, in part, to keep illustrations with reference to the unit-hypercube in order to provide a more intelligible demonstration of the mechanics of the *s-FLL* scheme. In section 4 the *s-FLL* scheme is described with reference to unit-hypercube. Section 5 puts the *s-FLL* scheme in perspective with selected neural-fuzzy and other models, and it delineates future extensions of our current work. The learning capacity of the *s-FLL* is demonstrated in section 6 where the *s-FLL* is applied for learning and recognition to two data sets in the Euclidean space and a third one involving jointly real numbers and lattice-ordered symbols. Eventually, in the conclusion-section 7, the main achievements are summarized.

2 FUZZY LATTICES AND A NOVEL LEARNING FRAMEWORK

In addition to the novel theoretical perspective for “automated learning” shown in this section we also include brief discussions regarding elementary lattice theory, for a reference purpose.

Recall that a *lattice* is a partly ordered set any two of whose elements have a greatest lower bound or “meet” denoted by $x\wedge y$, and a least upper bound or “join” denoted by $x\vee y$. Such an ordering is called herein *lattice ordering*. Regarding lattice theory the reader may refer to [1], [10], [38]. In this work we will be dealing primarily with *complete* lattices. A lattice is called *complete* when each of its subsets has a least upper bound and a greatest lower bound in the lattice in question. A nonvoid complete lattice contains a least element and a greatest element denoted herein by “O” and “I”, respectively, as suggested in [1], [38]. Note that sometimes the symbols “$\perp$” and “$\top$” are used in the literature to denote respectively the least and greatest elements of a complete lattice [10]. Some useful notational conventions are introduced next.

Typically a single capital letter is employed in the literature to denote a lattice [1], [10], [38]. Denoting a lattice by a single letter herein would simplify later notation but at the risk of confusing the reader. In particular, a risk stems from the fact that we consider two different lattice orderings of the same underlying set. Therefore in our notation for a lattice in this introductory paper we decided to discriminate an underlying set from a lattice ordering on its elements. Consequently we denote a lattice by an ordered pair $<\mathsf{A},\leq_\mathsf{A}>$, where A is the underlying set and $\leq_\mathsf{A}$ is a lattice ordering (binary) relation on the set A; that is $\leq_\mathsf{A}$ is a subset of $\mathsf{A}\times\mathsf{A}$ such that the *reflexive*, *antisymmetry*, and *transitive* laws hold and the above definition for a lattice applies. Two lattices sharing a common underlying set, say A, will be distinguished by the subscript of their ordering relations; for example $<\mathsf{A},\leq_{\mathsf{A},1}>$ and $<\mathsf{A},\leq_{\mathsf{A},2}>$. A lattice $<\mathsf{A},\leq_\mathsf{A}>$ will be called herein *conventional* lattice or *crisp* lattice. The *join* and *meet* operations in a crisp lattice $<\mathsf{A},\leq_\mathsf{A}>$ will be denoted respectively by $\vee_\mathsf{A}$ and $\wedge_\mathsf{A}$. Note that the subscript in all of $\leq_\mathsf{A}$, $\vee_\mathsf{A}$, and $\wedge_\mathsf{A}$, is meant to explicitly identify the lattice in question, whereas in the interest of simplicity we may omit using the subscript “A” for a lattice element. An exception to the latter rule is only the least and the greatest elements in a complete lattice $<\mathsf{A},\leq_\mathsf{A}>$ which will always be denoted respectively by O_A and I_A.

Instrumental to a simplified theoretical exposition of the work herein is the notion *dual* crisp lattice. The *dual* of a crisp lattice $<\mathsf{A},\leq_\mathsf{A}>$ has by definition the same underlying set but its partial ordering is the converse of $\leq_\mathsf{A}$. In this work we denote the dual of a lattice $<\mathsf{A},\leq_\mathsf{A}>$ by $<\mathsf{A},\leq^\partial_\mathsf{A}>$. Likewise, the *meet* and *join* operations in $<\mathsf{A},\leq^\partial_\mathsf{A}>$ are denoted respectively by $\wedge^\partial_\mathsf{A}$ and $\vee^\partial_\mathsf{A}$. Note that for $a,b\in<\mathsf{A},\leq_\mathsf{A}>$ it is both $a\vee_\mathsf{A}b=a\wedge^\partial_\mathsf{A}b$ and $a\wedge_\mathsf{A}b=a\vee^\partial_\mathsf{A}b$.

Another useful notion is the *product lattice* of two crisp lattices $<\mathsf{A},\leq_\mathsf{A}>$ and $<\mathsf{B},\leq_\mathsf{B}>$, that is in particular a crisp lattice defined on the Cartesian product $\mathsf{A}\times\mathsf{B}$ such that $(x_\mathsf{A},x_\mathsf{B})\leq_{\mathsf{A}\times\mathsf{B}}(y_\mathsf{A},y_\mathsf{B})$ if and only if $x_\mathsf{A}\leq_\mathsf{A}y_\mathsf{A}$ and $x_\mathsf{B}\leq_\mathsf{B}y_\mathsf{B}$ [1], [10], [38]. The *meet* in the product lattice $<\mathsf{A},\leq_\mathsf{A}>\times<\mathsf{B},\leq_\mathsf{B}>$ is given by $(x_\mathsf{A},x_\mathsf{B})\wedge(y_\mathsf{A},y_\mathsf{B})=(x_\mathsf{A}\wedge y_\mathsf{A},x_\mathsf{B}\wedge y_\mathsf{B})$, whereas the *join* is given by $(x_\mathsf{A},x_\mathsf{B})\vee(y_\mathsf{A},y_\mathsf{B})=(x_\mathsf{A}\vee y_\mathsf{A},x_\mathsf{B}\vee y_\mathsf{B})$ [1]. [10]. The product of N lattices ensues likewise. Apparently $<\mathsf{A},\leq_\mathsf{A}>\times<\mathsf{A},\leq_\mathsf{A}>$ is a product lattice, and it will be denoted by $<\mathsf{A}\times\mathsf{A},\leq_{\mathsf{A}\times\mathsf{A},1}>$ where the number *1* was used in the subscript so as to distinguish that lattice from a different one on $\mathsf{A}\times\mathsf{A}$ to be defined later in this section. Note that lattice $<\mathsf{A}\times\mathsf{A},\leq_{\mathsf{A}\times\mathsf{A},1}>$ is a *complete* one if $<\mathsf{A},\leq_\mathsf{A}>$ is a complete lattice. The least element in $<\mathsf{A}\times\mathsf{A},\leq_{\mathsf{A}\times\mathsf{A},1}>$ is $O_{\mathsf{A}\times\mathsf{A},1}=(O_\mathsf{A},O_\mathsf{A})$ and its greatest element is $I_{\mathsf{A}\times\mathsf{A},1}=(I_\mathsf{A},I_\mathsf{A})$.

Denoting succinctly a collection of lattice elements will be handy. Specifically assuming that I is an index set we denote a collection of elements of a lattice $<\mathsf{A},\leq_\mathsf{A}>$ by $\{a_i\}_{i\in I}$. When $<\mathsf{A},\leq_\mathsf{A}>$ is in particular a complete lattice then we denote the least upper bound of $\{a_i\}_{i\in I}$ by $\vee_\mathsf{A}\{a_i\}_{i\in I}$ and the greatest lower bound of $\{a_i\}_{i\in I}$ by $\wedge_\mathsf{A}\{a_i\}_{i\in I}$. Regarding additional notation, by writing $x\in<\mathsf{A},\leq_\mathsf{A}>$ we mean that x is an element of lattice’s underlying set A. When $x,y\in<\mathsf{A},\leq_\mathsf{A}>$ are related by $\leq_\mathsf{A}$ we write either $x\leq_\mathsf{A}y$ or $(x,y)\in\leq_\mathsf{A}$ For two elements x,y of a crisp lattice $<\mathsf{A},\leq_\mathsf{A}>$ note that either “x and y are *comparable*”, that is either $x\leq_\mathsf{A}y$ or $y\leq_\mathsf{A}x$, or “x and y are *incomparable*” that is neither $x\leq_\mathsf{A}y$ nor $y\leq_\mathsf{A}x$. In other words, when x and y are comparable either $(x,y)\in\leq_\mathsf{A}$ or $(y,x)\in\leq_\mathsf{A}$, whereas when x and y are incomparable both $(x,y)\notin\leq_\mathsf{A}$ and $(y,x)\notin\leq_\mathsf{A}$. The symbol $\|_\mathsf{A}$

denotes incomparability in lattice $<\mathsf{A},\leq_{\mathsf{A}}>$, in particular for incomparable lattice elements $x,y\in<\mathsf{A},\leq_{\mathsf{A}}>$ we write $x\|_{\mathsf{A}}y$.

An example of an all familiar lattice is $<\mathsf{R},\leq>$, where R is the set of real numbers. When $\leq$, $\vee$, $\wedge$, and $\|$ are used in this work without a subscript then they refer to the set of real numbers. The *meet* and *join* in $<\mathsf{R},\leq>$ are given respectively by $x\wedge y=min\{x,y\}$ and $x\vee y=max\{x,y\}$, where the operator *min* defines the minimum of its operands, and the operator *max* defines the maximum of its operands. Note that $x\|y$ is always false in $<\mathsf{R},\leq>$ and for this reason we call the set R *a chain* or, alternatively, *totally ordered set* [18].

The notion *fuzzy lattice* is introduced in order to extend, as explained in the sequel, the lattice ordering relation to all pairs (x,y) in the Cartesian product $\mathsf{A}\times\mathsf{A}$ including pairs of incomparable lattice elements. Such an extended relation may be regarded as a fuzzy set on the universe of discourse $\mathsf{A}\times\mathsf{A}$ [50]. In this work a fuzzy set will be denoted by (X,μ) where X is the universe of discourse and μ is a (membership) function μ: $\mathsf{X}\to[0,1]$. Hence, the aforementioned extended relation is denoted by $(\mathsf{A}\times\mathsf{A},\mu_R)$, and it is defined under the condition that it be 'compatible' with the crisp lattice's ordering relation $\leq_{\mathsf{A}}$ in the sense that $\mu_R(x,y)=1$ if and only if $x\leq_{\mathsf{A}}y$. Hence the crisp lattice ordering relation $\leq_{\mathsf{A}}$ corresponds to the fuzzy set $(\mathsf{A}\times\mathsf{A},\mu_L)$, where $\mu_L(x,y)=1$ if and only if $x\leq_{\mathsf{A}}y$, otherwise $\mu_L(x,y)=0$. Note that the fuzzy sets $(\mathsf{A}\times\mathsf{A},\mu_L)$ and $(\mathsf{A}\times\mathsf{A},\mu_R)$ are related by $(\mathsf{A}\times\mathsf{A},\mu_L)\subseteq(\mathsf{A}\times\mathsf{A},\mu_R)$ in the sense [48], that is $\mu_L(x,y)\leq\mu_R(x,y)$ for all $(x,y)\in\mathsf{A}\times\mathsf{A}$. The definition of a fuzzy lattice follows naturally.

Definition 1

A *fuzzy lattice* is a triplet $<\mathsf{A},\leq_{\mathsf{A}},\mu_R>$, where $<\mathsf{A},\leq_{\mathsf{A}}>$ is a crisp lattice and $(\mathsf{A}\times\mathsf{A},\mu_R)$ is a fuzzy set such that $\mu_R(x,y)=1$ if and only if $x\leq_{\mathsf{A}}y$.

We remark that if $<\mathsf{A},\leq_{\mathsf{A}},\mu_R>$ is a fuzzy lattice then its *dual* fuzzy lattice is defined as $<\mathsf{A},\leq^{\partial}_{\mathsf{A}},\mu^{\partial}_R>$ with $\mu^{\partial}_R(a,b)=\mu_R(b,a)$. We will refer to the collection of all fuzzy lattices by the term *framework of fuzzy lattices* or *FL-framework* for short. The significance of definition 1 is that it allows us to specify a degree of inclusion of a crisp lattice's element to any other, even when these two elements are incomparable. A further practical consequence is that via a fuzzy lattice we can deal with data other than vectors of real numbers; such data include symbols, propositions, etc.

Definition 1 stipulates only that the subset $(\mathsf{A}\times\mathsf{A},\mu_L)$ of the fuzzy set $(\mathsf{A}\times\mathsf{A},\mu_R)$ defines a crisp lattice ordering. No other restrictions apply whatsoever. Therefore $\mu_R(x,y)=1$ in a fuzzy lattice $<\mathsf{A},\leq_{\mathsf{A}},\mu_R>$ does not necessarily imply $\mu_R(y,x)=0$ and it could well be $\mu_R(y,x)>0$. Regarding transitivity in the fuzzy lattice $<\mathsf{A},\leq_{\mathsf{A}},\mu_R>$ we note that the conventional transitivity property holds only in the sense that $\mu_R(x,y)=1$ and $\mu_R(y,z)=1$ jointly imply $\mu_R(x,z)=1$. Whereas if, on the other hand, it is either $\mu_R(x,y)\neq1$ or $\mu_R(y,z)\neq1$ then $\mu_R(x,z)$ could be any number in [0,1].

The following definition for an *inclusion measure* will eventually enable the fuzzification of a crisp complete lattice.

Definition 2

An *inclusion measure* σ in a complete lattice $<\mathsf{A},\leq_{\mathsf{A}}>$ is a mapping σ: $<\mathsf{A},\leq_{\mathsf{A}}>\times<\mathsf{A},\leq_{\mathsf{A}}>\to[0,1]$ such that it satisfies the following conditions,

(C1) $\sigma(x,O_{\mathsf{A}})=0$, $x\neq O_{\mathsf{A}}$,

(C2) $\sigma(x,x)=1$, $\forall x\in<\mathsf{A},\leq_{\mathsf{A}}>$, and

(C3) $u\leq_{\mathsf{A}}w \Rightarrow \sigma(x,u)\leq\sigma(x,w)$, $u,w,x\in<\mathsf{A},\leq_{\mathsf{A}}>$ (Consistency Property).

where O_{A} is the least element in the complete lattice $<\mathsf{A},\leq_{\mathsf{A}}>$.

If I_{A} denotes the greatest element of a complete lattice $<\mathsf{A},\leq_{\mathsf{A}}>$ then it can be shown that $\sigma(x,I_{\mathsf{A}})=1$, $\forall\in<\mathsf{A},\leq_{\mathsf{A}}>$. The truth of the latter equation is shown as follows: it is $x\leq_{\mathsf{A}}I_{\mathsf{A}}$, $\forall\in<\mathsf{A},\leq_{\mathsf{A}}>$, and from conditions

(C2) and (C3) it can be inferred $x \leq_A I_A \Rightarrow \sigma(x,x)=1 \leq \sigma(x,I_A) \Rightarrow \sigma(x,I_A)=1$. The definition of the inclusion measure has been inspired from the notions "possibility measure" [49], [50] and "fuzzy measure" [25]. It can be argued that $\sigma(x,u)$ indicates the degree of inclusion of x in u. From henceforth the notations $\sigma(x,u)$ and $\sigma(x \leq_A u)$ will be employed interchangeably.

In the rest of this section we study sufficient conditions for the existence of an inclusion measure in a complete lattice. Along the way we also illustrate the context inside which an inclusion measure will be useful. The following notions will be needed.

Definition 3

Let P and Q be partly ordered sets, that is sets where the reflexive, antisymmetry, and transitive laws hold [10]. A mapping $\psi: P \to Q$ is said to be

(i) *order-preserving* (or, alternatively, *monotone*), if $x \leq y$ in P implies $\psi(x) \leq \psi(y)$ in Q;

(ii) an *order-isomorphism*, if both "$x \leq y$ in $P \Leftrightarrow \psi(x) \leq \psi(y)$ in Q", and "ψ is onto Q".

When there is an order-isomorphism from P to Q, we say that P and Q are order-isomorphic and we write $P \cong Q$. The same terminology applies to crisp lattices, hence an isomorphic relation between two crisp lattices will be denoted by $<\mathsf{A},\leq_A> \cong <\mathsf{B},\leq_B>$. Note that an "order-preserving" mapping had been called "isotone" in older texts on lattice theory [1], [38]; nevertheless herein we have employed instead the more modern term "monotone" [10] throughout.

Definition 4

Let $<\mathsf{A},\leq_A>$ and $<\mathsf{B},\leq_B>$ be lattices. A mapping $f: <\mathsf{A},\leq_A> \to <\mathsf{B},\leq_B>$ is said to be a (*lattice-*) *homomorphism* if f is *join-preserving* and *meet-preserving*, that is, for all $x,y \in <\mathsf{A},\leq_A>$, $f(x \vee y)=f(x) \vee f(y)$ and $f(x \wedge y)=f(x) \wedge f(y)$.

Note that a bijective (one-one) homomorphism is a (lattice-) isomorphism [10].

To pursue our quest for an inclusion measure in $<\mathsf{A},\leq_A>$ a real number will be attached to each of its elements by a *valuation* function. A valuation on a crisp lattice $<\mathsf{A},\leq_A>$ is a real-valued function $v: <\mathsf{A},\leq_A> \to <\mathsf{R},\leq>$ which satisfies $v(x)+v(y)=v(x \vee_A y)+v(x \wedge_A y)$, $x,y \in <\mathsf{A},\leq_A>$ [1], [38]. A valuation is *monotone* if and only if $x \leq_A y$ implies $v(x) \leq v(y)$, and *positive* if and only if $x <_A y$ implies $v(x) < v(y)$. It is known [1], [38] that a positive valuation v on a lattice $<\mathsf{A},\leq_A>$ renders the lattice in question a metric space with metric (distance) $d(x,y)=v(x \vee_A y)-v(x \wedge_A y)$, $x,y \in <\mathsf{A},\leq_A>$.

Consider the function $k(x,u)=v(u)/v(x \vee_A u)$, where $x,u \in <\mathsf{A},\leq_A>$, and v is a positive valuation on a complete lattice $<\mathsf{A},\leq_A>$ with $v(O_A)=0$. Note that if for a positive valuation v it is $v(O_A) \neq 0$, then another positive valuation v^+ with $v^+(O_A)=0$ can always be defined out of v by simply subtracting $v(O_A)$ from $v(x)$, $\forall x \in <\mathsf{A},\leq_A>$. We have assumed $v(O_A)=0$ for all the positive valuations considered in the sequel. The following theorem states that the function $k(x,u)$ defines an inclusion measure in $<\mathsf{A},\leq_A>$.

Theorem 5

The existence of a positive valuation function v on a complete lattice $<\mathsf{A},\leq_A>$, with $v(O_A)=0$, is a sufficient condition for the function $k(x,u)=v(u)/v(x \vee_A u)$ to be an inclusion measure in $<\mathsf{A},\leq_A>$.

The proof of theorem 5 is given in the Appendix.

We remark that $k(x,u)$ can be employed for showing the degree of inclusion of x in u, therefore in the sequel $k(x,u)$ will also be denoted by $k(x \leq_A u)$. Note that the function $k(x \leq_A u)$ equals 1 if and only if $x \leq_A u$. Therefore if v is a positive valuation function on a complete lattice $<\mathsf{A},\leq_A>$ then the triplet $<\mathsf{A},\leq_A,k>$ defines a fuzzy lattice. Theorem 5 shows one way for constructing an inclusion measure on a complete lattice $<\mathsf{A},\leq_A>$, and thus fuzzifying it, suffices the existence of a positive valuation in $<\mathsf{A},\leq_A>$. When more positive valuation functions are available in a specific application then the issue arises of selecting a suitable one.

Note that the selection of a suitable positive valuation function in a specific application can be regarded as a solution to an "optimization problem" for defining the function $\sigma=k$.

In the current work we have considered learning in the FL-framework as a procedure which handles (closed) lattice intervals as it is shown in section 4. One non-empty interval, say $[a,b]$ with $a,b\in$ <A,$\leq_A$> and $a\leq_A b$, is defined to be the non-empty set $[a,b]\triangleq\{x\in$ <A,$\leq_A$> : $a\leq_A x\leq_A b\}$. In the sequel we study the set of lattice intervals, then we propose 'constructions' of one or more intervals.

We, first, note that a single interval $[a,b]$ of <A,$\leq_A$> is a *sublattice* of <A,$\leq_A$>, where the notion 'sublattice' is defined next.

Definition 6

A lattice <S,$\leq_S$> is called *sublattice* of another lattice <A,$\leq_A$> if both $S\subseteq A$ and $a,b\in S \Rightarrow a\wedge_A b\in S$, $a\vee_A b\in S$.

The set VA of (closed) intervals in a complete lattice <A,$\leq_A$> augmented by the empty set, implies yet another crisp complete lattice as suggested by the following proposition.

Proposition 7

Let <A,$\leq_A$> be a complete lattice. Then the set VA of (closed) intervals in <A,$\leq_A$> augmented by the empty set O_{VA}, implies another complete lattice denoted by <VA,$\leq_{VA}$> with its least element denoted by O_{VA} $=[I_A,O_A]$. The implied lattice inclusion relation $[a,b]\leq_{VA}[c,d]$ in <VA,$\leq_{VA}$> is "$c\leq_A a$ and $b\leq_A d$". For two intervals $[a,b]$, $[c,d]\in$ <VA,$\leq_{VA}$>, their *lattice join* is given by $[a,b]\vee_{VA}[c,d]=[a\wedge_A c,b\vee_A d]$; and moreover their *lattice meet* is given by $[a,b]\wedge_{VA}[c,d]=[a\vee_A c,b\wedge_A d]$ if $a\vee_A c\leq_A b\wedge_A d$, otherwise it is $[a,b]\wedge_{VA}[c,d]=O_{VA}$.

The proof of proposition 7 is given in the Appendix.

In the FL-framework we have coined a term for referring to a finite collection of elements in a lattice as the following definition implies.

Definition 8

Let <A,$\leq_A$> be a lattice. A *family* (*of lattice* <A,$\leq_A$> *elements*) denoted by $f=\{w_i\}_{i\in I}$, is a finite collection of lattice <A,$\leq_A$> elements, that is $w_i\in$ <A,$\leq_A$> for all $i\in I$, and I is a finite index set.

The set of families corresponding to a lattice <A,$\leq_A$> will be denoted by F_A. Of particular interest in this work will be the set F_{VA} whose elements are families of lattice intervals. An interesting notion stems from considering the set-union of the members in a family of intervals. The terms *class* and *category* have been reserved and they are used interchangeably in the FL-framework to denote the set-union of the members in a family of intervals.

Definition 8

Let <A,$\leq_A$> be a lattice. A *class* (or, alternatively, *category*) *c in* <A,$\leq_A$> is defined as $c=\cup_{i\in I}w_i$, where $\{w_i\}_{i\in I}$ is a family in F_{VA}, and $\cup$ is the set-union operator.

The set of classes in a lattice <A,$\leq_A$> will be denoted by C_{VA}. Note that one class $c\in C_{VA}$ might have many decompositions into lattice intervals, that is many families of intervals may specify the same class. If for two different families of intervals $f_1=\{u_j\}_{j\in J}$ and $f_2=\{w_i\}_{i\in I}$ it holds $\cup_{j\in J}u_j=\cup_{i\in I}w_i$ then we say that both families of intervals f_1 and f_2 *represent the same class*; in other words $\{u_j\}_{j\in J}$ and $\{w_i\}_{i\in I}$ are two distinct decompositions of the same class. In this work we have restricted learning in lattices to the set F_{VA} rather than to the set C_{VA}, nevertheless we intend to employ the set C_{VA} in the future.

A *family activation function with respect to a concrete inclusion measure* σ *in* <VA,$\leq_{VA}$> is defined now.

Definition 10
A *family activation function with respect to a concrete inclusion measure* σ *in* <VA,$\leq_{VA}$> is defined to be a real valued function a_F: <VA,$\leq_{VA}$>×F_{VA}→[0,1], such that $a_F(x|f_c)= a_F(x|\{w_i\}_{i\in I})\triangleq max_{i\in I}\{\sigma(x\leq_{VA}w_i)\}$.

Note that a specific family $f_c\in F_{VA}$ implies a fuzzy set, that is the fuzzy set (VA,$a_F(x|f_c)$) on the universe of discourse VA. Given a $f_c\in F_{VA}$ the value $a_F(x_0|f_c)$ for a specific $x_0\in$<VA,$\leq_{VA}$> shows the degree of membership of x_0 in the fuzzy set (VA,$a_F(x|f_c)$) and it can be interpreted as the *activation* of a family f_c at the presence of x_0.

The previous analysis implies that an inclusion measure in the lattice <VA,$\leq_{VA}$> will be useful. Nevertheless, the only known way to the present authors for defining an inclusion measure is via a positive valuation. However there is no positive valuation in <VA,$\leq_{VA}$>, even when one exists in <A,$\leq_A$>. Further below, after the introduction of all the required mathematical tools we substantiate rigorously the claim that "a positive valuation in a complete lattice <A,$\leq_A$> can not imply another one in the complete lattice <VA,$\leq_{VA}$>".

The deficiency of an inclusion measure in <VA,$\leq_{VA}$> (starting from a positive valuation in <A,$\leq_A$>) was mended by introducing an injective monotone mapping of the lattice <VA,$\leq_{VA}$> to another lattice where an inclusion measure is available. To that end, (i) we assumed an isomorphism θ: <A,$\leq^{\partial}_A$>→<A,$\leq_A$>, and (ii) we introduced the *lattice of generalized* (or, alternatively, *abstract*) *intervals* <$A\times A$,$\leq_{A\times A,2}$>. Those issues are elaborated in the sequel.

Lemma 11
Let <A,$\leq_A$> be a complete lattice, then <$A\times A$,$\leq_{A\times A,2}$>=<A,$\leq^{\partial}_A$>×<A,$\leq_A$> is a complete lattice, namely the *lattice of generalized intervals* (or, alternatively, *lattice of abstract intervals*).

Note that the implied partial ordering relation in <$A\times A$,$\leq_{A\times A,2}$> is $[a,b]\leq_{A\times A,2}[c,d]$ if and only if $a\leq^{\partial}_A c \Leftrightarrow c\leq_A a$ and $b\leq_A d$, whereas the *meet* and *join* in lattice <$A\times A$,$\leq_{A\times A,2}$> are respectively $[a,b]\wedge_{A\times A,2}[c,d]=[a\wedge^{\partial}_A c,b\wedge_A d]= [a\vee_A c,b\wedge_A d]$ and $[a,b]\vee_{A\times A,2}[c,d]=[a\vee^{\partial}_A c,b\vee_A d]= [a\wedge_A c,b\vee_A d]$. Moreover the least and the greatest elements in the complete lattice <$A\times A$,$\leq_{A\times A,2}$> are given respectively by $[O^{\partial}_A,O_A]=[I_A,O_A]$ and $[I^{\partial}_A,I_A]=[O_A,I_A]$.

The aforementioned isomorphism θ: <A,$\leq^{\partial}_A$>→<A,$\leq_A$> implies that lattices <$A\times A$,$\leq_{A\times A,1}$>= <A,$\leq_A$>×<A,$\leq_A$> and <$A\times A$,$\leq_{A\times A,2}$>= <A,$\leq^{\partial}_A$>×<A,$\leq_A$> are isomorphic to each other. In particular assuming that Id_A is the identity mapping on A consider an isomorphic mapping θ×Id_A: <A,$\leq^{\partial}_A$>×<A,$\leq_A$>→ <A,$\leq_A$>×<A,$\leq_A$>given by $(\theta\times Id_A)(a,b)=(\theta(a),b)$. Hence in conclusion <$A\times A$,$\leq_{A\times A,1}$>≅<$A\times A$,$\leq_{A\times A,2}$>.

We study now the relation between the lattices <VA,$\leq_{VA}$> and <$A\times A$,$\leq_{A\times A,2}$>=. To this end we will need an injective monotone mapping i: <VA,$\leq_{VA}$>→<A,$\leq^{\partial}_A$>×<A,$\leq_A$> given by $i(\Delta)=[\wedge_A\Delta,\vee_A\Delta]$, where Δ∈<VA,$\leq_{VA}$>, $\wedge_A\Delta$ is the greatest lower bound of Δ and $\vee_A\Delta$ is the least upper bound of Δ in the complete lattice <A,$\leq_A$>.

We now conclude our quest for an inclusion measure in lattice <VA,$\leq_{VA}$>. Hence the composite injective monotone mapping $\phi=(\theta\times Id_A)\circ i$: <VA,$\leq_{VA}$>→<A,$\leq_A$>×<A,$\leq_A$> needs to be considered, where θ×Id_A and i have been given above. If we could find an inclusion measure σ in <A,$\leq_A$>×<A,$\leq_A$>=<$A\times A$,$\leq_{A\times A,1}$> then this same σ would be a valid inclusion measure in <VA,$\leq_{VA}$> due to the monotone mapping $\phi=(\theta\times Id_A)\circ i$. In other words $x,y\in$<VA,$\leq_{VA}$> implies that both their order-preserving images $\phi(x)$ and $\phi(y)$ are in <$A\times A$,$\leq_{A\times A,1}$> therefore it would make sense to calculate the number $\sigma(x\leq_{VA}y)= \sigma(\phi(x)\leq_{A\times A,1}\phi(y))$. It is easy to verify that such a σ in <VA,$\leq_{VA}$> satisfies all the conditions of definition 2; it suffices to consider the order-preserving

ϕ-images of the intervals $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle$. The following proposition 12 implies the existence of an inclusion measure in $\langle A\times A,\leq_{A\times A,1}\rangle$ in conjunction with Theorem 5.

Proposition 12
If v is a positive valuation on lattice $\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ then a function V: $\langle \mathsf{A},\leq_{\mathsf{A}}\rangle\times\langle \mathsf{A},\leq_{\mathsf{A}}\rangle\to\langle \mathsf{R},\leq\rangle$, given by $V(a,b)=v(a)+v(b)$, $a,b\in\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ defines a positive valuation on the lattice $\langle \mathsf{A},\leq_{\mathsf{A}}\rangle\times\langle \mathsf{A},\leq_{\mathsf{A}}\rangle=\langle A\times A,\leq_{A\times A,1}\rangle$.

The proof of proposition 12 is given in the Appendix.

To recapitulate, the objective had been the definition of an inclusion measure σ in the lattice $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle$ of intervals in order to make available a *family activation function* according to definition 10. The logical steps followed towards the aforementioned objective are enumerated underneath but not in the order given previously in the text.

1. A positive valuation v on $\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ implies a positive valuation V on $\langle A\times A,\leq_{A\times A,1}\rangle=\langle \mathsf{A},\leq_{\mathsf{A}}\rangle\times\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ by proposition 12, and hence it implies an inclusion measure σ on $\langle A\times A,\leq_{A\times A,1}\rangle$ by theorem 5.
2. The lattice $\langle A\times A,\leq_{A\times A,2}\rangle=\langle \mathsf{A},\leq^{\partial}_{\mathsf{A}}\rangle\times\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ was defined by lemma 11.
3. We considered an isomorphism θ: $\langle \mathsf{A},\leq^{\partial}_{\mathsf{A}}\rangle\to\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$.
4. A mapping $\theta\times Id_{\mathsf{A}}$: $\langle A\times A,\leq_{A\times A,2}\rangle\to\langle A\times A,\leq_{A\times A,1}\rangle$, where Id_{A} is the identity mapping on A, implied the isomorphism $\langle A\times A,\leq_{A\times A,1}\rangle\cong\langle A\times A,\leq_{A\times A,2}\rangle$.
5. We considered an injective monotone mapping i: $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle\to\langle A\times A,\leq_{A\times A,2}\rangle$.
6. We considered the composite injective monotone mapping $\phi=(\theta\times Id_{\mathsf{A}})\circ i$: $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle\to\langle A\times A,\leq_{A\times A,1}\rangle$. Hence ϕ maps x,y from $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle$ to their order-preserving images $\phi(x)$, $\phi(y)$ in $\langle A\times A,\leq_{A\times A,1}\rangle$.
7. In conclusion an inclusion measure $\sigma(x\leq_{\mathsf{VA}}y)$ became available in lattice $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle$, with $\sigma(x\leq_{\mathsf{VA}}y)= k(\phi(x)\leq_{\mathsf{A\times A},1}\phi(y))$ where k is defined by theorem 5.

Our conclusion is summarized in the following Lemma.

Lemma 13
Let $\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ be a complete lattice with a positive valuation v: $\langle \mathsf{A},\leq_{\mathsf{A}}\rangle\times\langle \mathsf{A},\leq_{\mathsf{A}}\rangle\to\langle \mathsf{R},\leq\rangle$, and let θ: $\langle \mathsf{A},\leq^{\partial}_{\mathsf{A}}\rangle\to \langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ be an isomorphism. Then the function k: $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle\times\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle\to[0,1]$ given by $k([a,b],[c,d])= [v(\theta(c))+v(d)]/[v(\theta(a\wedge_{\mathsf{A}}c))+v(b\vee_{\mathsf{A}}d)]$ defines an inclusion measure in the complete lattice $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle$ of intervals.

Note that lattices $\langle A\times A,\leq_{A\times A,1}\rangle$ and $\langle A\times A,\leq_{A\times A,2}\rangle$ have been both defined on the same underlying set $\mathsf{A}\times\mathsf{A}$ but each one considers a different ordering of the elements of $\mathsf{A}\times\mathsf{A}$. Thus lattice $\langle A\times A,\leq_{A\times A,1}\rangle$ orders the elements of $\mathsf{A}\times\mathsf{A}$ by the relation $\leq_{A\times A,1}$, whereas lattice $\langle A\times A,\leq_{A\times A,2}\rangle$ orders the elements of the same set $\mathsf{A}\times\mathsf{A}$ by the relation $\leq_{A\times A,2}$. To distinguish the elements of these two lattices we have made a convention. That is an element of lattice $\langle A\times A,\leq_{A\times A,1}\rangle$ is embraced by parentheses as in (a,b) with $a,b\in\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$, whereas an element of lattice $\langle A\times A,\leq_{A\times A,2}\rangle$ is embraced by brackets as in $[a,b]$ with $a,b\in\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$.

It is interesting to note that the injective monotone mapping i is a join-preserving but not a meet-preserving mapping. To see that i is join-preserving recall that the join of $[a,b]$ and $[c,d]$ is given by $[a\wedge_{\mathsf{A}}c,b\vee_{\mathsf{A}}d]$ in both lattices $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle$ and $\langle A\times A,\leq_{A\times A,2}\rangle$. To see that i is not meet-preserving apply it to the meet of two different one-point intervals.

At this point we substantiate our previous claim that "a positive valuation in $\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ can not imply another one in $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle$". Hence consider the composite mapping $V\circ(\theta\times Id_{\mathsf{A}})\circ i$: $\langle \mathsf{VA},\leq_{\mathsf{VA}}\rangle\to\langle \mathsf{R},\leq\rangle$, where V is the positive valuation on $\langle A\times A,\leq_{A\times A,1}\rangle=\langle \mathsf{A},\leq_{\mathsf{A}}\rangle\times\langle \mathsf{A},\leq_{\mathsf{A}}\rangle$ shown in proposition 12, and $\phi=(\theta\times Id_{\mathsf{A}})\circ i$ has been defined previously. Due to the fact that the mapping i is not meet-preserving we conclude that $V\circ(\theta\times Id_{\mathsf{A}})\circ i$ is not a valuation on $<\mathsf{VA},\leq_{\mathsf{VA}}>$, let alone a positive valuation.

All the analysis in this section was carried out regarding a lattice $<\mathsf{A},\leq_{\mathsf{A}}>$. However, it has to be pointed out that $<\mathsf{A},\leq_{\mathsf{A}}>$ could itself be the product of N lattices, called *constituent lattices*. In such a case and in line with the notation already introduced we write $<\mathsf{A},\leq_{\mathsf{A}}>=<\mathsf{A}_1,\leq_{\mathsf{A}1}>\times\ldots\times<\mathsf{A}_{\mathrm{N}},\leq_{\mathsf{AN}}>$. If all the lattices $<\mathsf{A}_1,\leq_{\mathsf{A}1}>,\ldots,<\mathsf{A}_{\mathrm{N}},\leq_{\mathsf{AN}}>$ are *complete* with respective least/greatest elements $O_{\mathsf{A}1}/I_{\mathsf{A}1},\ldots,O_{\mathsf{AN}}/I_{\mathsf{AN}}$ then $<\mathsf{A},\leq_{\mathsf{A}}>$ is a complete lattice with least element $(O_{\mathsf{A}1},\ldots,O_{\mathsf{AN}})$ and greatest element $(I_{\mathsf{A}1},\ldots,I_{\mathsf{AN}})$ [1].

Proposition 14
Let $v_1,\ldots,v_{\mathrm{N}}$ be valuations on lattices $<\mathsf{A}_1,\leq_{\mathsf{A}1}>,\ldots,<\mathsf{A}_{\mathrm{N}},\leq_{\mathsf{AN}}>$, respectively. Then a function v: $<\mathsf{A}_1,\leq_{\mathsf{A}1}>\times\ldots\times<\mathsf{A}_{\mathrm{N}},\leq_{\mathsf{AN}}>\rightarrow<\mathsf{R},\leq>$ given by $v(x_1,\ldots,x_{\mathrm{N}})=v_1(x_1)+\ldots+v_{\mathrm{N}}(x_{\mathrm{N}})$, where $x_{\mathrm{i}}\in<\mathsf{A}_{\mathrm{i}},\leq_{\mathsf{Ai}}>$, i=1,…,N, defines a valuation on $<\mathsf{A},\leq_{\mathsf{A}}>=<\mathsf{A}_1,\leq_{\mathsf{A}1}>\times\ldots\times<\mathsf{A}_{\mathrm{N}},\leq_{\mathsf{AN}}>$.

The proof is analogous to the proof of proposition 12. Note that it suffices to be all the valuations $v_1,\ldots,v_{\mathrm{N}}$ monotone so as the valuation v to be monotone as well. If at least one of the monotone valuations $v_1,\ldots,v_{\mathrm{N}}$ is, in addition, a positive valuation then v is a positive valuation.

To simplify notation we have made a convention regarding the elements of lattices $<A\times A,\leq_{A\times A,1}>$ and $<A\times A,\leq_{A\times A,2}>$ when $<\mathsf{A},\leq_{\mathsf{A}}>$ is a product lattice. Then we denote an element of lattice $<A\times A,\leq_{A\times A,1}>$ by $(a_1,b_1,\ldots,a_{\mathrm{N}},b_{\mathrm{N}})$, where $(a_1,b_1)\in<A_1\times A_1,\leq_{A1\times A1,1}>,\ldots,(a_{\mathrm{N}},b_{\mathrm{N}})\in<A_{\mathrm{N}}\times A_{\mathrm{N}},\leq_{AN\times AN,1}>$; whereas we denote an element of lattice $<A\times A,\leq_{A\times A,2}>$ by $[a_1,b_1,\ldots,a_{\mathrm{N}},b_{\mathrm{N}}]$, where $[a_1,b_1]\in<A_1\times A_1,\leq_{A1\times A1,2}>,\ldots,[a_{\mathrm{N}},b_{\mathrm{N}}]\in<A_{\mathrm{N}}\times A_{\mathrm{N}},\leq_{AN\times AN,2}>$.

3 A COMPLETE LATTICE IN THE UNIT-HYPERCUBE

In the previous section, aspects of the framework of fuzzy lattices or FL-framework have been detailed. The theory presented is valid for any mathematical lattice suffices the lattice in question is a complete one and two functions are available: (i) a positive valuation function v: $<\mathsf{A},\leq_{\mathsf{A}}>\rightarrow<\mathsf{R},\leq>$, where R is the set of real numbers, and (ii) an isomorphic function θ: $<\mathsf{A},\leq^{\partial}_{\mathsf{A}}>\rightarrow<\mathsf{A},\leq_{\mathsf{A}}>$.

In this section a complete lattice will be studied equipped with the aforementioned functions. The lattice in question is defined on the N-dimensional unit-hypercube U, with partial ordering $(x_1,\ldots,x_{\mathrm{N}})\leq_{\mathsf{U}}(y_1,\ldots,y_{\mathrm{N}})\Leftrightarrow x_1\leq y_1,\ldots,x_{\mathrm{N}}\leq y_{\mathrm{N}}$, where $(x_1,\ldots,x_{\mathrm{N}})$, $(y_1,\ldots,y_{\mathrm{N}})\in\mathsf{U}$. In the interest of simplicity the dimension of the unit-hypercube will not be cited explicitly therefore the corresponding lattice will be denoted by $<\mathsf{U},\leq_{\mathsf{U}}>$.

It is important to note that dealing with the unit-hypercube U instead of dealing with the whole set R^{N} is a valid assumption in “machine learning” [28] and it does not impair by default the learning capacity since all physical quantities have upper and lower bounds, and suitable transformations to the unit hypercube can be found. An interval in lattice $<\mathsf{U},\leq_{\mathsf{U}}>$ is an N-dimensional hyperbox, or *hyperbox* for short, in the unit-hypercube. Recall that the learning objective is to identify sets of points in $<\mathsf{U},\leq_{\mathsf{U}}>$ which (sets) can be represented as families of hyperboxes, the latter are finite collections of hyperboxes.

Lattice $<\mathsf{U},\leq_{\mathsf{U}}>$ is the product of N identical constituent lattices, these are the *chains* $<\mathtt{I},\leq>$, where $\mathtt{I}$=[0,1] and ≤ is the conventional *less-than or equal-to* operator between real numbers. Recall that a *chain* implies a *total ordering* relation of its elements [18], that is $x\|_{\mathtt{I}}y$ is false for all x,y in $<\mathtt{I},\leq>$. Moreover note that each of the N (lattice) chains $<\mathtt{I},\leq>$, is a *complete* one with least element $O_{\mathtt{I}}$=0.0 and greatest element $I_{\mathtt{I}}$=1.0.

By proposition 14 in the previous section and the comments underneath it, we infer that if the v_{i}’s, $i\in\{1,\ldots,N\}$ are monotone valuations on the constituent (lattice) chains $<\mathtt{I},\leq>$ then the function

$v(x_1,\ldots,x_N)=v_1(x_1)+\ldots+v_N(x_N)$ defines an monotone valuation on $<\mathsf{U},\leq_{\mathsf{U}}>$. If at least one of v_i, $i\in\{1,\ldots,N\}$ is, in addition, a positive valuation then $v(x_1,\ldots,x_N)=v_1(x_1)+\ldots+v_N(x_N)$ is a positive valuation. Therefore the problem of searching for a positive valuation on lattice $<\mathsf{U},\leq_{\mathsf{U}}>$ has been reduced to the problem of finding a positive valuation on a chain $<\mathtt{I},\leq>$. But the latter problem is trivial. Specifically any monotone increasing function on the unit interval [0,1] is a positive valuation on the chain $<\mathtt{I},\leq>$. In this paper, linear positive valuations of the form $v_i(x_i)=c_ix_i$, $i\in\{1,\ldots,N\}$ with $c_i>0$ have been considered, resulting in a positive valuation of the form $v(x_1,\ldots,x_N)=c_1x_1+\ldots+c_Nx_N$, where $c_i>0$, $i\in\{1,\ldots,N\}$ on lattice $<\mathsf{U},\leq_{\mathsf{U}}>$.

On the other hand, the isomorphism $\theta_{\mathtt{I}}(x)=1-x$, $x\in[0,1]$ was employed in lattice $<\mathtt{I},\leq>$. In turn $\theta_{\mathtt{I}}$ implies an isomorphism, denoted by θ, in the product lattice $<\mathsf{U},\leq_{\mathsf{U}}>$ as follows : $\theta((a_1,\ldots,a_N))=(\theta_{\mathtt{I}}(a_1),\ldots,\theta_{\mathtt{I}}(a_N))=(1-a_1,\ldots,1-a_N)$. Thus an element $(a,b)=((a_1,\ldots,a_N),(b_1,\ldots,b_N))=(a_1,b_1,\ldots,a_N,b_N)$ of lattice $<\mathsf{U}\times\mathsf{U},\leq_{\mathsf{U}\times\mathsf{U},1}>$ is mapped to its isomorphic element $[\theta(a),b]=[\theta(a_1,\ldots,a_N),(b_1,\ldots,b_N)]=[(1-a_1,\ldots,1-a_N),(b_1,\ldots,b_N)]=[1-a_1,b_1,\ldots,1-a_N,b_N]$ of lattice $<\mathsf{U}\times\mathsf{U},\leq_{\mathsf{U}\times\mathsf{U},2}>$. And vice-versa, an element $[a,b]=[(a_1,\ldots,a_N),(b_1,\ldots,b_N)]=[a_1,b_1,\ldots,a_N,b_N]$ of lattice $<\mathsf{U}\times\mathsf{U},\leq_{\mathsf{U}\times\mathsf{U},2}>$ is mapped to its isomorphic element $(\theta(a),b)=(\theta(a_1,\ldots,a_N),(b_1,\ldots,b_N))=((1-a_1,\ldots,1-a_N),(b_1,\ldots,b_N))=(1-a_1,b_1,\ldots,1-a_N,b_N)$ of lattice $<\mathsf{U}\times\mathsf{U},\leq_{\mathsf{U}\times\mathsf{U},1}>$.

Figure 1 illustrates some of the notions introduced in the previous section with reference to the lattice $<\mathsf{U},\leq_{\mathsf{U}}>$ for $N=2$, that is the unit-square on the plane. Note that a rectangle (a box) corresponds to an interval in lattice $<\mathsf{U},\leq_{\mathsf{U}}>$. Figures 1 (a) and (b) demonstrate the partial ordering relation. In Fig.1(a) the points (a_1,a_2) and (b_1,b_2) are comparable, in particular it is $(a_1,a_2)\leq(b_1,b_2)$. Whereas in Fig.1(b) it is neither $(c_1,c_2)\leq(d_1,d_2)$ nor $(d_1,d_2)\leq(c_1,c_2)$, that is the points (c_1,c_2) and (d_1,d_2) are incomparable symbolically $(c_1,c_2)\|_{\mathsf{U}}(d_1,d_2)$. Fig.1(c) shows how two different families (of boxes) f_1 and f_2 inside the unit-square may be defined as collections of boxes; in the particular case of Fig.1(c), overlapping boxes were put in the same family but it could be otherwise. Finally Fig.1(d) illustrates the significance of the consistency property (C3, in definition 2), where box x is specified by the product of intervals [.15,.2]×[.15,.2], box u is specified by [.5,.6]×[.3,.4], and box w is specified by [.4,.9]×[.2,.8]. Thus

$\sigma(x\leq u)=\ v(u)/v(x\vee u)=\ v(.5,.6,.7,.4)/v((.85,.2,.85,.2)\vee(.5,.6,.7,.4))=\ v(.5,.6,.7,.4)/v(.85,.6,.85,.4)=\ 2.2/2.7\approx 0.8148$, and

$\sigma(x\leq w)=\ v(w)/v(x\vee w)\}=\ v(.6,.9,.8,.8)/v((.85,.2,.85,.2)\vee(.6,.9,.8,.8))=\ v(.6,.9,.8,.8)/v(.85,.9,.85,.8)=3.1/3.4\approx 0.9118$.

That is, as guaranteed by the consistency property, the degree of inclusion of box x in box w is larger than the degree of inclusion of x in u, because box u is inside box w. A definition useful in the next section regards the *size* of a hyperbox and is given underneath.

Definition 15

Let $<\mathsf{VA},\leq_{\mathsf{VA}}>$ be the lattice of intervals stemming from a lattice $<\mathsf{A},\leq_{\mathsf{A}}>$, and let v be a positive valuation on $<\mathsf{A},\leq_{\mathsf{A}}>$. Then a function Z: $<\mathsf{VA},\leq_{\mathsf{VA}}>\rightarrow<\mathsf{R},\leq>$, given by $Z([a,b])=v(b)-v(a)$, defines the *size of an interval* $[a,b]$ *with respect to the positive valuation* v.

Note that if $t_r=[a,b]$ is a trivial lattice interval with $a=b$ then it follows $Z(t_r)=Z([a,a])=v(a)-v(a)=0$, that is the size of a trivial interval is zero. Furthermore note that the size of the least element $[I_{\mathsf{A}},O_{\mathsf{A}}]$ in VA is $Z([I_{\mathsf{A}},O_{\mathsf{A}}])=v(O_{\mathsf{A}})-v(I_{\mathsf{A}})=-v(I_{\mathsf{A}})$, since it has been assumed $v(O_{\mathsf{A}})=0$. When definition 15 is applied in particular to lattice $<\mathsf{U},\leq_{\mathsf{U}}>$ with positive valuation $v(x_1,\ldots,x_N)=x_1+\ldots+x_N$ then the size $Z([a,b])$ of a hyperbox $[a,b]=[(a_1,\ldots,a_N),(b_1,\ldots,b_N)]=[a_1,b_1]\times\ldots\times[a_N,b_N]$ is given by $Z([a,b])=v(b)-v(a)=v(b_1,\ldots,b_N)-v(a_1,\ldots,a_N)=\Sigma(b_i-a_i)$. In this specific case the size of the least element $[I_{\mathsf{U}},O_{\mathsf{U}}]$ in lattice $<\mathsf{VU},\leq_{\mathsf{VU}}>$ equals $Z([I_{\mathsf{U}},O_{\mathsf{U}}])=-v(I_{\mathsf{U}})=-N$.

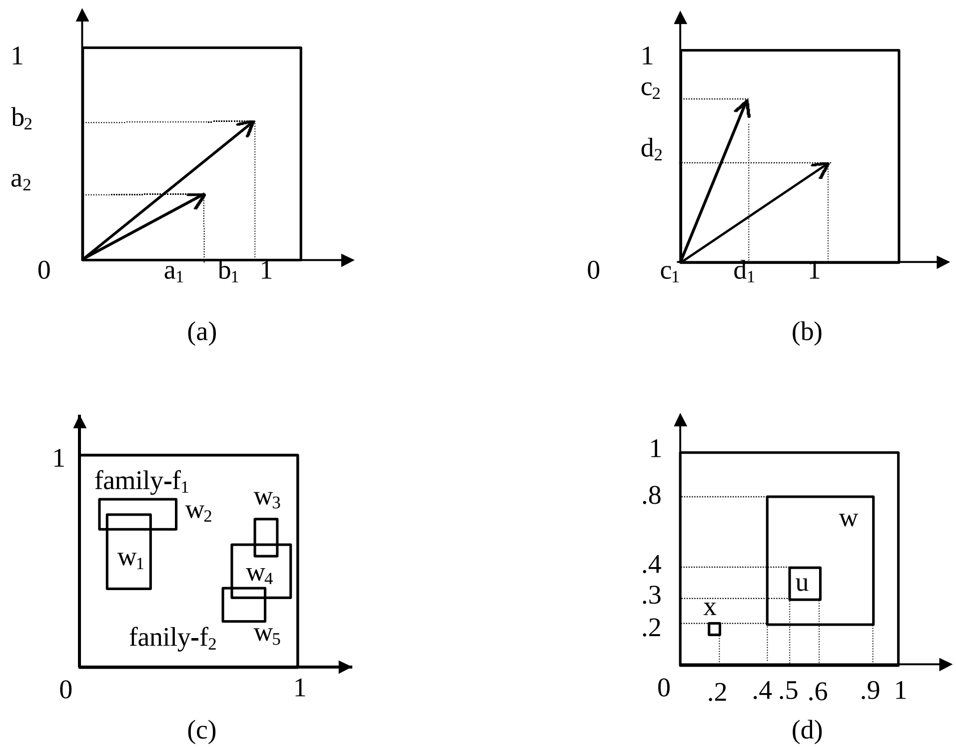

Figure 1 Interpretations in the 2-dimensional lattice $(\mathsf{U},\le_{\mathsf{U}})$, that is the unit-square.

(a) Two comparable elements of lattice $(\mathsf{U},\le_{\mathsf{U}})$ with $(a_1,a_2) \le_{\mathsf{U}} (b_1,b_2)$.

(b) The elements (c_1,c_2) and (d_1,d_2) of lattice $(\mathsf{U},\le_{\mathsf{U}})$ are incomparable. Symbolically $(c_1,c_2)\|_{\mathsf{U}}(d_1,d_2)$.

(c) The families f_1 and f_2 are defined as collections of overlapping boxes.

(d) The consistency property guarantees that when one box u is inside another box w then a box x is included in w more than it is in u.

4 σ - FUZZY LATTICE LEARNING (*s-FLL*) SCHEME

In this section the σ- fuzzy lattice learning scheme (*s-FLL* scheme) is presented with reference to the lattice $<\mathsf{U},\le_{\mathsf{U}}>$. Even though the *s-FLL* scheme is applicable to a general complete lattice $<\mathsf{A},\le_{\mathsf{A}}>$ equipped with (i) a positive valuation function *v*: $<\mathsf{A},\le_{\mathsf{A}}>\to<\mathsf{R},\le>$, where R is the set of real numbers, and (ii) an isomorphic function θ: $<\mathsf{A},\le^{\partial}_{\mathsf{A}}>\to<\mathsf{A},\le_{\mathsf{A}}>$, we decided to introduce the *s-FLL* scheme in the context of lattice $<\mathsf{U},\le_{\mathsf{U}}>$ for two reasons. First, because the acquaintance with the Euclidean space is expected to help a reader to understand easier the mechanics of the *s-FLL* scheme. Second, because since the *s-FLL* scheme has been inspired from the (fuzzy) adaptive resonance theory (ART) [9] which is applicable solely in the Euclidean space, an introduction of the *s-FLL* scheme in the unit-hypercube U is expected to help those readers who are familiar with the fuzzy ART to trace similarities with the *s-FLL*. Traits of the *s-FLL* scheme are discussed in this section, whereas an explicit but brief comparison with the fuzzy ART is outlined in section 5. A detailed comparison with the fuzzy ART is a topic for a future publication.

The goal of learning by the *s-FLL* is to locate clusters in the data by hyperboxes, these are elements of lattice $<\mathsf{VU},\le_{\mathsf{VU}}>$ of intervals/hyperboxes in $<\mathsf{U},\le_{\mathsf{U}}>$. A hyperbox defines a set of akin, in the sense of adjacent, elements in lattice $<\mathsf{U},\leq_{\mathsf{U}}>$. Let $f_k=\{w_{k,i}\}_{i\in Ik}$, $k=1,\ldots,M<\infty$ be families of hyperboxes. A hyperbox in a family f_k is called a *constituent hyperbox*. Only *simplified* families $\{w_{k,i}\}_{i\in Ik}$ will be considered which are defined as follows.

Definition 16

A family of hyperboxes $f_k=\{w_{k,i}\}_{i\in Ik}$ is called *simplified* when it has no "nested" hyperboxes, that is when there are no two constituent hyperboxes $w_{k,m}$, $w_{k,n} \in <\mathsf{VU},\leq_{\mathsf{VU}}>$ in f_k such that $w_{k,m}\leq_{\mathsf{VU}}w_{k,n}$.

The *s-FLL* for learning by clustering is described next.

4.1 The σ -FLL Scheme for Clustering

0. The first input $x_0\in<\mathsf{VU},\leq_{\mathsf{VU}}>$ is memorized. From then on learning and decision making proceed as follows (assume that at any instant t there are $M(t)<\infty$ known and learned families - that is $M(t)<\infty$ finite collections of hyperboxes are kept in the memory).
1. Present an input $x_c\in<\mathsf{VU},\leq_{\mathsf{VU}}>$ to the initially "set" database of families $f_1,\ldots,f_{M(t)}$.
2. While there exist families f_k, $k=1,\ldots,M(t)<\infty$ that have not been "reset", calculate the activation $a_F(x_c|f_k)$ of the "non-reset" f_k's at the presence of x_c.
3. Competition among the f_k's : select the f_J which is activated the most at the presence of x_c, where J is the index of the winner family, and let w_L denote the hyperbox of $f_J=\{w_{J,i}\}_{i\in IJ}$ which corresponds to the $max_{i\in IJ}\{\sigma(x_c\leq_{\mathsf{VA}}w_{J,i})\}$.
4. The *Assimilation Condition*: Test whether the size $Z(x_c\vee_{\mathsf{U}}w_L)$ - with respect to the selected positive valuation - of $x_c\vee_{\mathsf{U}}w_L$ is less than a user defined threshold size Z_{crit}, where w_L was defined in the previous step.
5. If the assimilation condition is satisfied, then replace w_L in f_J by $x_c\vee_{\mathsf{U}}w_L$. In the sequel simplify, if necessary, families by eliminating "nested" constituent hyperboxes in a family.
6. If the assimilation condition fails then "reset" f_J.
7. If all the families $f_1,\ldots,f_{M(t)}$ have been "reset" and no winner has been found then memorize hyperbox x_c provided there is sufficient system memory available. When no memory is available turn down the accommodation of the input x_c.

In order to make this section self-contained and to facilitate an implementation of the *s-FLL* scheme by other researchers we have appended a list of all the needed functions after the "*s-FLL* for Testing" algorithm in the next subsection.

In the previous scheme, learning is continuously "on" for all presented inputs. In particular, a new input $x_c\in<\mathsf{VU},\leq_{\mathsf{VU}}>$ is assigned to the most activated family provided that the corresponding size $Z(x_c\vee_{\mathsf{U}}w_L)$ (step 4) is under a user defined threshold Z_{crit}. Otherwise "reset" occurs and the quest for a new winner resumes. If all the families $f_1,\ldots,f_{M(t)}$ have been "reset" and no winner has been found then the input x_c is memorized. 'Memorization' means that an input x_c is learned as a new cluster. Moreover note that a simplification is carried out in step 5 in order to remove clusters which are redundant in view of the consistency property (C3, in definition 2).

A reader familiar with the fuzzy ART [9] might have already noticed the analogies with the *s-FLL* scheme during the learning of hyperboxes. Certain analogies are pointed out explicitly in section 5. Nevertheless, a fundamental difference with the fuzzy ART is that the *s-FLL* is applicable on a lattice domain (including the unit-hypercube lattice $<\mathsf{U},\leq_{\mathsf{U}}>$) due to its lattice applicable inclusion measure σ.

Despite *s-FLL*'s capacity to keep learning (by clustering) continuously "on", there exist practical cases like these in the examples of section 6, where learning is expected by classification. Then one data set is given for training and another data set is given for testing. In such a case the aforementioned algorithm is used for *training* in order to effect learning. In the sequel, learning is disengaged and the *s-FLL* scheme for testing is employed as described in the sequel.

4.2 The σ -FLL Scheme for Testing

1. The next input $x_c \in <\mathsf{VU}, \leq_{\mathsf{VU}}>$ shows up.
2. Calculate the activation $a_F(x_c|f_k)$ of each learned family f_k, $k=1,\ldots,M_F<\infty$ at the presence of x_c. That is calculate the degree of membership of $x_c \in <\mathsf{VU}, \leq_{\mathsf{VU}}>$ to all the fuzzy sets (VU, $a_F(x|f_k)$), $k=1,\ldots,M_F<\infty$.
3. Assign x_c to the winner of the competition among all families. That is assign x_c to the family f_k, $k=1,\ldots,M_F<\infty$ which provides with the largest value (VU, $a_F(x|f_k)$) of membership of x_c in it.

Note that during testing the total number of families does not change with time but instead it equals a constant positive integer value $M_F<\infty$.

At this point we cite explicitly expressions of the functions needed by the *s-FLL* scheme, during both "learning" and "testing".

- A hyperbox x_c in the N-dimensional unit-hypercube U is denoted by $x_c=[a,b]=[(a_1,\ldots,a_N),(b_1,\ldots,b_N)]=[a_1,b_1,\ldots,a_N,b_N]$, where $a_i,b_i \in [0,1]$ and $[a_i,b_i]$ specifies an interval in the i^{th} dimension for $i=1,\ldots,N$.
- Let $w=[p,q]=[(p_1,\ldots,p_N),(q_1,\ldots,q_N)]=[p_1,q_1,\ldots,p_N,q_N]$ denote another hyperbox.
- $x_c \vee_{\mathsf{U}} w$ is calculated by $x_c \vee_{\mathsf{U}} w= [a_1,b_1,\ldots,a_N,b_N] \vee_{\mathsf{U}} [p_1,q_1,\ldots,p_N,q_N]= [min\{a_1,p_1\},max\{b_1,q_1\},\ldots, min\{a_N,p_N\},max\{b_N,q_N\}]$.
- $Z(x_c)$ is calculated by $Z(x_c)=Z([a_1,b_1,\ldots,a_N,b_N])=\Sigma(b_i-a_i)$.
- We have considered the positive valuations $v_i(x)=x$ and the isomorphisms $\theta_i(x)=1-x$, $x\in[0,1]$ along the i^{th} dimension for $i=1,\ldots,N$. Hence the number $v(\theta(p))+v(q)$ is calculated as $v(\theta(p))+v(q)=v(\theta(p_1,\ldots,p_N))+v(q_1,\ldots,q_N)=v(1-p_1,\ldots,1-p_N)+ v(q_1,\ldots,q_N)=\Sigma(1-p_i)+\Sigma q_i=N+\Sigma(q_i-p_i)$.
- The degree of inclusion of hyperbox x_c in hyperbox w is calculated by $\sigma(x_c \leq_{\mathsf{U}} w)= [v(\theta(p))+v(q)]/[v(\theta(a \wedge_{\mathsf{U}} p))+v(b \vee_{\mathsf{U}} q)]= [N+\Sigma(q_i-p_i)]/[\mathrm{N}+\Sigma(max\{b_i,q_i\}-min\{a_i,p_i\})]$.
- A family $f_k=\{\mathrm{w}_{k,i}\}_{i\in Ik}$ of hyperboxes comprises "$|I_k|$ many" hyperboxes where $|I_k|$ denotes the (finite) cardinal number of the set I_k.
- For a concrete family f_k, the number $a_F(x_c|f_k)$ is calculated as $a_F(x_c|f_k)=a_F(x_c|\{\mathrm{w}_{k,i}\}_{i\in Ik})= max_{i\in Ik}\{\sigma(x_c \leq w_{k,i})\}$.

In all, the *s-FLL* scheme is a competitive learning scheme for a crisp set identification which employs a lattice inclusion measure σ in the lattice $<\mathsf{VU}, \leq_{\mathsf{VU}}>$ of hyperboxes. Learning by the *s-FLL* scheme occurs by clustering and a hyperbox defines a cluster. The total number of hyperboxes (clusters) to be learned is not known *a priori* by the *s-FLL* scheme, but rather this number is specified on-line during learning. Moreover, the total number of learned hyperboxes may be partitioned into $M(t)<\infty$ families (during learning) or $M_F<\infty$ families (during testing) either by an external teacher who labels the learned hyperboxes or automatically. There are several ways for automatic partitioning. One way is to assume that a family contains only a single hyperbox; another way is to assume that two overlapping hyperboxes are in the same family; yet another way is to assume that two hyperboxes are in the same family only when the degree of inclusion of one hyperbox into the other is over a user defined threshold. Despite the fact that the hyperboxes which specify each one of the $M(t)$ families (during training) or the M_F families (during testing) are crisp, generalization is still possible by the activation function $a_F(x|f_k)$ which can specify the degree of membership of a hyperbox $x_c \in <\mathsf{VU}, \leq_{\mathsf{VU}}>$ in a fuzzy set (VU, $a_F(x|f_k)$).

An advantage of the *s-FLL* scheme compared to the fuzzy ART is that its inputs do not have to be solely points but they could be hyperboxes as well. Another advantage shared with the fuzzy ART is the reduced training complexity; specifically a single pass through the data suffices for stable learning, in the sense that additional passes through the data will not alter the arrangement of the hyperboxes in the learned families. Note that learning by "a single pass" through the data is an inherent characteristic of the *s-FLL* scheme. However the price we have to pay for such a rapid learning is that the families of hyperboxes being formed during learning depend on the presentation order of the training data. Specifically, a different presentation order of the training data causes, in general, different formations of hyperboxes and consequently it occasions different families of hyperboxes. In turn different families, say f_1 and f_2 with $f_1 \neq f_2$, give rise to different activations functions $a_F(x|f_1)$ and $a_F(x|f_2)$. Hence, eventually, the "direction" of learning is altered by selecting a different winner among the competing families of hyperboxes. We haven't studied the dependence of classification performance on the order of the training data; such a study for the *s-FLL* is a topic of future research, and a similar study recently published on the fuzzy ART [17] could be the starting point. Finally note that the pre-eminent advantage of the *s-FLL* scheme is its applicability beyond the Euclidean space as well on products of complete lattices, provided that both a positive valuation v and an isomorphism θ exist in each constituent lattice. One example demonstrating *s-FLL*'s applicability beyond the Euclidean space is shown in example 6.3 of section 6.

5 RELATED WORK IN PERSPECTIVE

The *s-FLL* scheme may be regarded as a fuzzy neural network (FNN). Connections to selected FNN(s) as well as to other models are drawn in this section aiming at putting the *s-FLL* in perspective. However due to the emphasis of the present paper on the mathematical substantiation of the FL-framework, a detailed comparison with any specific model is outside the scope of this paper.

Recall that a FNN combines the "bottom-up" neural capacity for parallel learning and generalization, with fuzzy logic's "top-down" capacity for human-like reasoning based on uncertain or ill defined data. The list of FNN(s) cited below was not meant to be comprehensive. Rather, we have picked out only a few models some aspects of which are so closely related to our work that a comparison would be instructive.

The *s-FLL* scheme has stemmed from a study of the adaptive resonance theory (ART) [3], [6], [7], [9], in particular the fuzzy-ART [9]. Some common traits are summarized next: (i) Learning by the *s-FLL* scheme corresponds to fuzzy ART's "fast learning", (ii) Fuzzy ART's "complement coding" corresponds to *s-FLL*'s isomorphic function θ, (iii) Both the fuzzy ART and the *s-FLL* imply a two layer architecture; the one (input) layer accommodates the inputs, the other (category) layer accommodates the learned intervals/hyperboxes, (iv) Search for a "winner-take-all" occurs in the category layer. When the winner does not satisfy *s-FLL*'s assimilation condition (or ART's vigilance criterion) then this winner is "reset" and search keeps on, (v) Fuzzy ART's vigilance parameter (ρ) corresponds to *s-FLL*'s threshold size (Z_{crit}) which is used for regulating the formation of hyperboxes/clusters during learning, and (vi) The inclusion measure (σ) of the *s-FLL* corresponds to both of fuzzy ART's "choice function" and "match function" as it will be detailed in a future publication.

Apart from the fuzzy ART itself there exist other FNN(s) that employ the fuzzy ART as one of their constituent modules. For instance FALCON [26] employs the fuzzy ART synergistically with the backpropagation. Note that all FNN(s) which employ the fuzzy ART can potentially enhance their applicability beyond the Euclidean space to a mathematical lattice domain by considering the *s-FLL* scheme. Examples of lattices wherein the *s-FLL* is applicable are cited further below in this section.

Regarding other FNN(s) we note SONFIN [24], that is a neural fuzzy inference network with on-line capacity for self-construction. The FNN in [34] provides fuzzy interpretations to the multilayer perceptron (MLP) and it can deal with inputs in numerical, linguistic, and set form by representing *N*-dimensional patterns as *3N*-dimensional vectors. The FNN in [31] builds on [34] but the emphasis is on an "expert" connectionist model, and the network generates a measure of certainty expressing confidence in its decisions. The knowledge-based network in [30] builds on [31], [34] and it improves performance by encoding initial knowledge in terms of class *a priori* probabilities; the output to an input interval is computed by using only the mean of an interval. Intervals of vectors are treated explicitly by either supervised of unsupervised Min-Max neural networks [39], [40]. A neural architecture for handling fuzzy input vectors is proposed in [20] then a learning algorithm is derived for propagating fuzzy numbers through the neural architecture; such a propagation is effected by the computationally expensive task of dealing with *h*-level sets. In [11] six different methods are shown for fuzzy modeling by rapid rule extraction without *a priori* information about the data. Inference in the fuzzy logic-based neural network (FLBN) [27] occurs by neuron-implemented fuzzy AND/OR logic rules, and antecedents are weighted by a possibility measure.

The majority of the FNN(s) employ either explicitly or implicitly the *N*-dimensional Euclidean space, that is the product lattice $<\mathsf{R},\leq>^N$ where R is the set of real numbers. In particular, the majority of the FNN(s) deal solely with propositions involving values of linguistic variables defined over the line of the real numbers. Hence only such fuzzy sets are usually dealt with as “small”, “large”, “fast”, “old”, etc. where the universe of discourse is *totally ordered* and in particular it is the real line. In other words most FNN(s) do not deal directly with linguistic variables involving *colors* such as “blue, red, yellow”, *properties* such as “edible, poisonous, dexterous”, *gender* such as “masculine, feminine”, etc. unless the user maps arbitrarily the values of the corresponding linguistic variables to the *N*-dimensional space $<\mathsf{R},\leq>^N$ for a suitable value of *N*. The *s-FLL* provides the alternative for dealing with a *lattice ordered* universe of discourse, for instance lattice-ordered symbols, as demonstrated in example 6.3 of section 6.

We remark that “mapping the data to a lattice” compared to “mapping the data to the real-line” could enhance a subsequent learning scheme’s capacity for data discrimination. To be more specific, due to the partial ordering relation implied by a lattice compared to the total ordering relation of the real-line, the present authors expect that two or more classes to be potentially easier separable on a lattice than on the real line. The same remark extends to more lattices/real-lines. Hence the capacity to treat rigorously and jointly disparate lattice elements by the competitive, self-organizing *s-FLL* scheme is expected to be significant in decision making. Moreover “*s-FLL* learning” is characterized by a single pass through the training set, but the cost for such a rapid learning is “dependence on the order of the training data” as it has been explained at the end of section 4. Another factor for enhancing potentially *s-FLL*’s data processing speed is its treatment of one lattice interval as a single *block* in the computations, since the *s-FLL* operates on the lattice of intervals.

There is only one parameter (one number) to tune during learning by the *s-FLL* herein, that is number $Z_{crit}\in[0,1]$, which specifies the maximum allowable size of the hyperboxes/clusters to be learned. In this sense we may say that the value of Z_{crit} regulates “learning by the *s-FLL*” the same way as the learning-rate parameter regulates “learning by the multilayer perceptron (MLP)”; whereas the numbers defining the edges of a hyperbox in the *s-FLL* scheme correspond to the weights of a MLP. Nevertheless unlike the MLP, learning by the *s-FLL* is always stable and only a single pass through the data is enough. Note that larger or smaller values of Z_{crit} effect learning reversely. On the one hand, “larger values of Z_{crit}” result in fewer hyperboxes/clusters while the capacity for data discrimination is deteriorated, whereas on the other hand “smaller values of Z_{crit}” result in more hyperboxes/clusters with an enhanced capacity for data discrimination.

The *s-FLL* can deal with “uncertainty” in the data by treating whole intervals of data values instead of treating only individual data points. Alternative ways for dealing with uncertainty in the data include considering the lattice of fuzzy sets on the universe of discourse, or the lattice of events in a probability space. These issues will be detailed in future publications.

The crisp intervals being identified during learning by the *s-FLL* can be thought of as encoding knowledge regarding the outside world. The way it has been introduced in this work, the *s-FLL* scheme sets out its learning without *a priori* knowledge. However, initial *a priori* knowledge can be provided to the *s-FLL* by letting a user specify a number of lattice intervals before “*s-FLL* learning” takes place due to the training data.

The *s-FLL* scheme can also be regarded as a single layer fuzzy logic inference system that consists solely of lattice applicable IF-THEN rules. The presentation of the *s-FLL* as a lattice applicable fuzzy inference system, including consideration of an additional layer of OR gates, is a topic of future publications.

Note also that *s-FLL*'s inclusion measure σ is an alternative to fuzzy set theory's "possibility measure". Recall that an inclusion measure σ can be calibrated by a positive valuation function *v*. In turn, *s-FLL*'s positive valuation function *v* can be regarded as a mathematical sound alternative to either the *probability measure* of Bayesian statistics or the *belief functions* of the Dempster-Schaffer theory. Finally we underline the potential utility of the inclusion measure σ for system identification in modeling and control [21]. For instance σ could be employed by a "reasoning scheme" in a diversified environment involving disparate types of lattice-ordered data aiming at the selection of a suitable system model [42].

6 EXAMPLES

The *s-FLL* scheme has been applied for pattern recognition to three data sets, including two benchmarks. Two of those data sets involved solely real numbers. The third data set, that is a benchmark, intermingled real numbers and lattice-ordered symbols. In all cases during learning the formed lattice intervals (clusters) were labeled by an external teacher; then a testing set was employed to evaluate the capacity of the *s-FLL* for generalization. For the two benchmark data sets, where one training set and one testing set are given explicitly, we report the percentages of correct classification on the testing set by the *s-FLL* scheme comparatively with other learning schemes. We have assumed that the "other" studies whose results are reported herein have employed the same training and testing data sets. Therefore we conclude that a comparison of the performance by the various algorithms reported herein is meaningful and fair.

6.1 Raman-Spectroscopy data set

The task is to identify the soft tissues encountered in the epidural surgical procedure [35] using their Raman spectra profiles [41]. 27 Raman spectra profiles were available that corresponded to four soft tissues encountered in the epidural surgical procedure. The data profiles were given by an expert separated into four groups; six data corresponded to connective-tissue, seven data to muscle-tissue, six to skin-tissue, and another eight data corresponded to fat-tissue. Figure 2 shows the emission intensity signatures of four Raman spectra profiles, that is one typical representative for each one of the four soft tissues encountered in the epidural surgical procedure. Some of the data of each group were selected randomly and formed the training set, and the remaining data of the group were used as testing data.

The discriminatory power of the Fourier Transform's phase [33] was employed. In the case of Raman spectra it was found that only the first 22 phase components were significantly greater than zero, therefore all the rest of the phase components were discarded. Each vector was normalized by a linear transformation which mapped the minimum value of a vector entry to the number 0.0 and the maximum value of a vector entry to the number 1.0. The lattice where the *s-FLL* scheme applied was $<\mathsf{U},\leq_{\mathsf{U}}>$ in the 22 dimensions. The positive valuation function employed was the sum of all the entries of the corresponding normalized vector.

A series of learning experiments for soft tissue recognition was carried out. Various combinations from 6 to 14 data profiles in the aggregate were selected randomly from each one of the four soft tissue groups and formed the training data set. The remaining data were employed as the testing set. Every time, and as a result of training, four distinct clusters were specified corresponding to each one of the "connective", "muscle", "skin", and "fat" tissues. An external teacher attached labels to the learned clusters, those clusters were 22-dimensional hyperboxes. In the sequel the testing data were applied. No misclassifications have been recorded.

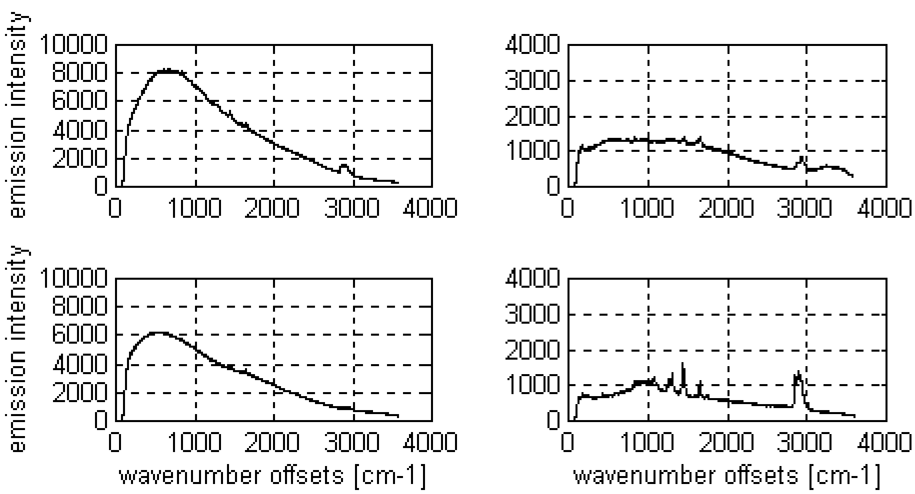

Figure 2 Raman spectra profiles showing laser emission intensities versus wavenumber offsets away from the driving frequency at 633 nm that correspond to four soft tissues encountered in the epidural surgical procedure.

6.2 VOWEL benchmark data set

This data set was taken from the Carnegie Mellon University's (CMU) collection of neural net benchmarks [13]. It consists of 10-dimensional vectors of the linear predictive coefficients with regards to eleven steady state vowels of British English. The training set contains 528 vectors, while the testing set consists of 462 vectors. The aim is the correct classification of the testing data to their corresponding classes, these are the eleven steady state vowels. Table 1 shows the classification results by several different methods. The entries of Table 1 have been inserted in a reduced order of performance; that is a method achieving a larger "percent correct" on the testing data was put above another one which achieved a smaller "percent correct". The authors experimented only with the *s-FLL* scheme which has been applied in lattice $<U,\leq_U>$ in the 10 dimensions.

The best results shown in Table 1 by the three-dimensional Growing Cell Structures, that is 67%, are reported in [14] where as many as 80 training epochs had to be carried out in order to achieve the best result. The best results by the Gaussian ARTMAP and by the Fuzzy ARTMAP, either without voting or with 5 voters, are reported in [47]. A common feature shared by all the three : Gaussian ARTMAP, Fuzzy ARTMAP, and *s-FLL*, is their rapid training. The rest of the entries in Table 1 have been reported by other researchers and relevant information can be found in the documentation which accompanies this benchmark [13].

TABLE 1
Performance of various methods in classifying eleven vowels of British English of the VOWEL benchmark. The methods are arrayed in a decreasing order of success.

Method	No. of Hidden Units	% Correct
3-D Growing Cell Structures	154	67
Gaussian ARTMAP (5 voters)	273	63
σ-FLL Scheme	195	60.17
Gaussian ARTMAP (w/o voting)	55	59
Nearest Neighbor	-	56.27
Square Node Network	88	54.76

Details of *s-FLL*'s performance for learning the VOWEL benchmark are shown in Table 2. Several different types of data normalization have been tried. A row in Table 2 corresponds to one normalization type. For each normalization type eight different ranges for the threshold size Z_{crit} were tested. These are Range-1: from 0 to 0.018 (0-0.018) ,…, Range-8: from 0.14 to 0.158 (0.14-0.158) shown in Table 2. A column in Table 2 corresponds to one range. Within a range, say Range-4 (0.06-0.078), the size Z_{crit} has been increased by a step of 0.002. A cell of Table 2 displays the maximum (M) and minimum (m) recorded "percentage correct" of the testing test for a normalization type and a range.

On the one hand note that normalization of the data has been, in effect, an additional preprocessing of the data. On the other hand, recall that the size Z_{crit} is the only parameter of the *s-FLL* scheme that needs to be defined by the user in order to regulate the learning as it has been illustrated in section 5.

The first normalization type ((c1) in Table 2) was achieved by finding the minimum and the maximum entry values for each one of the 10 entries of the training input vectors, and those values were mapped to the numbers 0.0 and 1.0 respectively. The best recorded testing performance was 52.81%. For cases (c2) through (c4) of Table 2 three different normalization types were tried. In particular an input vector's entry was normalized respectively over the normalizing intervals [-6,6], [-8,8], and [- 10,10]. In all those cases the positive valuation function employed was $v(x_1,\ldots,x_{10})=x_1+\ldots+x_{10}$. Finally, cases (c5) through (c7) in Table 2 were treated respectively by the same normalizing intervals as cases (c2) through (c4) above, nevertheless the positive valuation function was altered by employing the standard deviation of each input vector entry. In particular instead of using the function $v(x_1,\ldots,x_{10})=x_1+\ldots+x_{10}$, where $(x_1,\ldots,x_{10})$ is a normalized input vector, we used the function $v_{new}(x_1,\ldots,x_{10})=\sigma_1 x_1+\ldots+\sigma_{10} x_{10}$, where σ_i, i=1,…,10 is the standard deviation of the corresponding vector entry in the training set. Employment of that new positive valuation function was motivated by the hypothesis that entries which are more closely concentrated around their means could convey a lesser discriminatory power than other, more dispersed entries. The experimental testing verified this hypothesis and the performance improved. Specifically in case (c7) and Range-6 the *s-FLL* scheme achieved a maximum of 60.17 %, and as shown in Table 1, 195 cluster/hyperboxes were located in the data. An important advantage of learning by the *s-FLL* has been "a single pass" through the training data set, that is an inherent property of learning by the *s-FLL*.

Regarding the behavior of *s-FLL*'s performance during the experiments, we note that *s-FLL*'s performance was fairly stable for one type of data normalization. More specifically, within one type of data normalization no drastic variation of performance was noted over the size's Z_{crit} "whole range of change" as shown across a line in Table 2. Finally, it is noted that the total number of clusters/hyperboxes learned in the experiments was found to be a decreasing function of the size Z_{crit}. For instance the numbers of clusters/hyperboxes learned for sizes Z_{crit} = 0.005, 0.055, 0.105, 0.155 and normalization type (c7) were respectively 527, 344, 217, 162.

TABLE 2

Details of *s-FLL*'s performance for learning the VOWEL benchmark. Seven combinations of data normalization and positive valuations (c1)…(c7) are shown across the rows of Table 2, for eight different ranges of the size Z_{crit} shown along the columns of Table 2. In row (c1) a data entry was normalized over its minimum-maximum range. In rows (c2) through (c4) a data entry was normalized respectively over the ranges [-6,6], [-8,8], and [-10,10]. The same normalization intervals were used in rows (c5) through (c7), where the asterisk (*) is to show that new positive valuation functions $v_{new}(.)$ have been used as explained in the text. In a range corresponding to a column, say Range-1 (0-.018), the size Z_{crit} was increased in steps of 0.002. A cell of Table 2 displays the maximum (M) and the minimum (m) "percentage correct" of the testing data.

Type of data Normalization	Maximum (M) and minimum (m) Percentage of Correct Classification in 8 Ranges for the size of Z_{crit}							
	Range-1 0 - .018	Range-2 .02 - .038	Range-3 .04 - .058	Range-4 .06 - .078	Range-5 .08 - .098	Range-6 .10 - .118	Range-7 .12 - .138	Range-8 .14 - .158
(c1) Full range	50.65 M 50.65 m	50.65 M 50.65 m	50.65 M 50.65 m	50.65 M 50.65 m	50.65 M 50.65 m	50.87 M 50.65 m	51.95 M 51.08 m	52.81M 52.38 m
(c2) [-6,6]	55.84 M 55.84 m	55.84 M 55.63 m	56.28 M 55.63 m	56.28 M 54.55 m	58.87 M 57.58 m	58.44 M 55.63 m	55.84 M 52.81 m	56.28 M 53.03 m
(c3) [-8,8]	55.84 M 55.84 m	56.49 M 55.63 m	56.06 M 54.33 m	58.87 M 56.71 m	56.49 M 54.11 m	56.49 M 52.81 m	56.93 M 54.76 m	57.58 M 55.84 m
(c4) [-10,10]	55.84 M 55.84 m	56.49 M 55.41 m	58.66 M 54.55 m	58.01 M 54.98 m	56.49 M 53.03 m	57.36 M 55.19 m	56.06 M 52.60 m	53.46 M 51.06 m
(c5) [-6,6] *	57.79 M 57.79 m	57.79 M 57.79 m	58.44 M 57.79 m	57.58 M 55.84 m	58.23 M 57.36 m	58.66 M 57.14 m	58.01 M 56.06 m	59.09 M 56.71 m
(c6) [-8,8] *	57.79 M 57.79 m	58.44 M 57.79 m	58.23 M 56.06 m	58.66 M 57.14 m	58.44 M 55.63 m	58.66 M 56.28 m	59.09 M 56.71 m	59.52 M 58.44 m
(c7) [-10,10] *	57.79 M 57.79 m	58.44 M 57.58 m	58.44 M 55.63 m	58.44 M 56.49 m	58.87 M 56.49 m	60.17 M 57.14 m	58.87 M 56.06 m	57.79 M 54.55 m

6.3 ABALONE benchmark data set

This data comes from the University of California Irvine (UCI) collection of benchmark data sets [29] and it includes information which could be readily encoded in a lattice outside the Euclidean space. The data are gender and physical measurements on abalones, the latter are marine snails. There exist 4177 records of data, in particular 3133 for training and 1044 for testing. Each data record has 9 entries. The first 8 entries are one gender plus seven physical measurements data and they are used as inputs. The ninth entry indicates one of three age groups and it is used as an output. The question is to learn predicting the age of an abalone.

Unique to the ABALONE benchmark is the first input entry. It contains information regarding the gender of an abalone and it receives three values these are '*I*', '*M*', '*F*' that stand for *I*nfant, *M*ale, and *F*emale. Other researchers have encoded the first input entry as three inputs to a neural network [46]. Hence to reach the best performance in Table 3 the values '*I*', '*M*', '*F*' were encoded by the vectors (1,0,0), (0,1,0), and (0,0,1) respectively. Note that such an encoding has been decided arbitrarily by a user and it employs implicitly a Boolean lattice of features, while on the other hand the said encoding increases the total number of inputs from 8 to 10. The authors employed the product of two lattices. The first lattice, denoted by $<\mathsf{L},\leq_{\mathsf{L}}>$, encoded the first ABALONE input entry as a single input to the *s-FLL* by considering the lattice of Figure 3, hence $\mathsf{L}=\{I,M,F,G\}$ and $I\leq_{\mathsf{L}}F$, $I\leq_{\mathsf{L}}M$, $F\leq_{\mathsf{L}}G$, $M\leq_{\mathsf{L}}G$ as shown in Fig.3. The second lattice was $<\mathsf{U},\leq_{\mathsf{U}}>$ in the 7 dimensions to encode the remaining input entries.

Regarding the first lattice the symbols '*F*' and '*M*' were treated as (lattice) incomparable, that is $F\|_{\mathsf{L}}M$, whereas both of them were treated as "greater", in a lattice sense, than the symbol '*I*'. The symbol '*G*' in Fig.3 was introduced by the authors in order to make the lattice in question a complete one so as the *s-FLL* scheme be applicable. Figure 3 shows also the values of the employed positive valuation function. The isomorphic function θ employed has been $\theta(I)=G$, $\theta(F)=F$, $\theta(M)=M$, and $\theta(G)=I$.

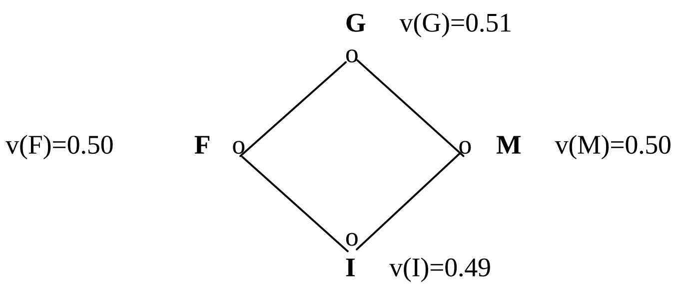

Figure 3 The three symbols '*I*', '*F*', '*M*' in the first input entry of the ABALONE benchmark were considered to be lattice ordered. Symbol '*G*' was introduced by the authors to make the corresponding lattice $<\mathsf{L},\leq_{\mathsf{L}}>$ a complete one. The values of the employed positive valuation are also shown.

Table 3 shows the classification results by different methods. The entries of Table 3 are inserted in a reduced order of performance. The authors experimented only with the *s-FLL* scheme. The remaining entries in Table 3 are reported with additional information in the documentation which accompanies this benchmark [29]. The performance of the *s-FLL* scheme was marginally the second best but without introducing the additional overhead of increasing the number of inputs as explained above. Moreover only 1 pass through the training data sufficed for stable learning, as expected from the *s-FLL* scheme.

TABLE 3

The *s-FLL* scheme achieved marginally the second best performance. Due to its treatment of the ABALONE data as lattice elements, the *s-FLL* scheme reduced the total number of inputs.

Method	% Correct
Cascade-Correlation (5 hidden nodes)	65.61
σ-FLL scheme	65.32

Back-Propagation	64.00
Nearest Neighbour (k=5)	62.46
Cascade-Correlation (no hidden nodes)	61.40
C4.5	59.20
Dystal	55.00
Linear Discriminant Analysis	32.57

During learning a teacher attached labels to the learned intervals which (labels) corresponded to the majority of the data encoded by a lattice interval. One category consisted of a number of lattice intervals (clusters). The total number of clusters was a decreasing function of the value of the threshold size Z_{crit}. We note that the sizes Z_{crit}= 0.500, 0.600, and 0.700 yielded respectively 346, 256, and 192 clusters. In our experiments the size of Z_{crit} was increased from 0.500 to 1.00 in steps of 0.001. The average classification correct rate on the testing data was 60.65 % with standard deviation 1.24. The performance peaked to 65.32 % as reported in Table 3 for threshold size Z_{crit}=0.636 without normalizing the numeric data. Moreover in the peak of the performance 223 clusters were located in the data.

Additional experiments have been carried out. In some of them we verified that the performance has been a function of the training data order presentation, as expected from the *s-FLL* scheme. The best results were recorded for the ordering in which the data are given in the UCI collection. Moreover, in other experiments, the ABALONE data were employed with different lattice orderings and with different positive valuations of the first input entry. The best results were recorded for the lattice ordering and positive valuation shown in Fig.3. Note that in all the additional experiments the performance deteriorated by up to 3 percentage points.

In conclusion, this example has shown the efficient applicability of the *s-FLL* scheme for learning and recognition of lattice-ordered symbols in conjunction with real numbers. Specifically, the first input entries of the ABALONE data set have been treated as lattice-ordered symbols; we did not have to convert those symbols to real numbers. Moreover, the *s-FLL* scheme scored high as shown in Table 3 and its learning had been rapid. An additional advantage of the *s-FLL* scheme is that its encoding, using lattice ordered symbols, did not increase the total number of inputs.

7 CONCLUSION

The principal contribution of this paper has been the introduction of a novel base for learning, that is the framework of fuzzy lattices, or FL-framework for short. New notions were introduced in the context of the FL-framework such as the notion of a fuzzy lattice as well as the notion of an inclusion measure σ. An inclusion measure σ can fuzzify the crisp binary inclusion relation in a conventional complete lattice. In the FL-framework, learning can be achieved by handling crisp lattice intervals. Various learning schemes can be synthesized in the FL-framework. One such a scheme for learning by clustering has been introduced herein, that is the σ- fuzzy lattice learning (*s-FLL*) scheme, inspired from the fuzzy adaptive resonance theory (ART) [9]. For convenience of comparison with the fuzzy ART the *s-FLL* has been introduced in the unit-hypercube U. The *s-FLL* scheme has been also presented in perspective with selected fuzzy/neural/expert systems from the literature. The capacity of the *s-FLL* has been demonstrated on three data sets. Solely real numbers were treated in a medical data set and the VOWEL benchmark, whereas real numbers mixed with lattice-ordered symbols were treated in the ABALONE benchmark.

Advantages of both the *s-FLL* scheme and the FL-framework include, (i) the capacity to deal with the uncertainty in the data in several alternative manners including treatment of “lattice intervals” these are “hyperboxes” in the Euclidean unit-hypercube U, (ii) the ability for sound, hierarchic, and modular decision making in a diversified (lattice) environment, based (the decision making) on an inclusion measure σ.

Regarding the *s-FLL* scheme in particular, its employment in specific learning experiments has demonstrated its inherent capacity for rapid learning; in particular only 1 pass through the data suffices for stable learning. Nevertheless the order of data presentation is significant for reasons exposed herein. Finally note that the *s-FLL* scheme herein requires only one parameter (one number) to be defined by the user in order to carry out learning, the parameter in question is the size threshold Z_{crit}.

Future plans include a study and a further development of various theoretical aspects of the FL-framework, as those plans have been delineated throughout this paper and mainly in section 5. A summary of the said plans includes a suitable selection of a positive valuation v for a concrete application, an optimum selection of an isomorphic function θ in a lattice, a procedure that would maximize the degree of inclusion of a class into another one, etc. Moreover and in addition to learning by clustering, other learning schemes will be investigated in the FL-framework.

ACKNOWLEDGMENTS

The authors are grateful to the anonymous reviewers for initial encouragement as well as for constructive suggestions and comments. Furthermore, we would like to acknowledge the contribution of the Raman spectra data by Peter Brett, Tim Parker from the AMARC, Department of Mechanical Engineering, University of Bristol, UK, and John Day from the ``Interface Analysis Centre'', Bristol, UK. The kind assistance of Sam Waugh from the University of Tasmania for conveying promptly his valuable experience with the ABALONE data set is also acknowledged.

APPENDIX

Proof of Theorem 5

The proof underneath is valid for any complete lattice $<\mathsf{A},\leq_\mathsf{A}>$ with a positive valuation v and $v(O_\mathsf{A})=0$.

1. $k(x,O_\mathsf{A})= v(O_\mathsf{A})/v(x\vee_\mathsf{A}O_\mathsf{A})= v(O_\mathsf{A})/v(x)=0$, assuming $x\neq O_\mathsf{A}$.

Note that $k(x,O_\mathsf{A})$ is guaranteed to be equal to 0 (for $x\neq O_\mathsf{A}$) only if v is a positive, and not just a monotone, valuation.

2. If $x=O_\mathsf{A}$ then we define $k(O_\mathsf{A},O_\mathsf{A})=1$, otherwise it holds $k(x,x)= v(x)/v(x\vee_\mathsf{A}x)= v(x)/v(x)=1$.

3. In any lattice $<\mathsf{A},\leq_\mathsf{A}>$ the operation of join ($\vee_\mathsf{A}$) is monotone [1], that is $u\leq_\mathsf{A}w \Rightarrow x\vee_\mathsf{A}u\leq_\mathsf{A}x\vee_\mathsf{A}w$.

Provided a positive, and hence a monotone, valuation v in lattice $<\mathsf{A},\leq_\mathsf{A}>$ a distance function is defined in $<\mathsf{A},\leq_\mathsf{A}>$ by $d(u,w)=v(u\vee_\mathsf{A}w)-v(u\wedge_\mathsf{A}w)$ [1], [38]. Which, in conjunction with the above monotone inequality, implies both $d(x\vee_\mathsf{A}u,x\vee_\mathsf{A}w)=v(x\vee_\mathsf{A}w)-v(x\vee_\mathsf{A}u)$ and $d(u,w)=v(w)-v(u)$. In what follows it will be convenient to remind that we put no subscript to ≤ when it is used with real numbers.

It is further known from [1] that if v is a monotone valuation in $<\mathsf{A},\leq_\mathsf{A}>$ then it holds $d(x\vee_\mathsf{A}u,x\vee_\mathsf{A}w)+ d(x\wedge_\mathsf{A}u,x\wedge_\mathsf{A}w)\leq. d(u,w)$. Hence in our case with an initial assumption $u\leq_\mathsf{A}w$, it follows
$d(x\vee_\mathsf{A}u,x\vee_\mathsf{A}w) \leq d(x\vee_\mathsf{A}u,x\vee_\mathsf{A}w)+d(x\wedge_\mathsf{A}u,x\wedge_\mathsf{A}w)\leq. d(u,w) \Rightarrow v(x\vee_\mathsf{A}w)-v(x\vee_\mathsf{A}u) \leq v(w)-v(u) \Rightarrow v(x\vee_\mathsf{A}w) \leq v(w)-v(u)+v(x\vee_\mathsf{A}u)$.

Assuming $w\neq O_\mathsf{A}$ it follows that
$[v(u)/v(w)]v(x\vee_\mathsf{A}w) \leq [v(u)/v(w)][v(w)-v(u)+v(x\vee_\mathsf{A}u)] = [(v(w)-v(u))/v(w)]v(u)+[v(u)/v(w)]v(x\vee_\mathsf{A}u) \leq [(v(w)-v(u))/v(w)]v(x\vee_\mathsf{A}u)+[v(u)/v(w)]v(x\vee_\mathsf{A}u) = v(x\vee_\mathsf{A}u) \Rightarrow [v(u)/v(x\vee_\mathsf{A}u)]\leq[v(w)/v(x\vee_\mathsf{A}w)] \Rightarrow k(x,u) \leq k(x,w)$.

Proof of Proposition 7

The truth of this proposition will be shown in four steps.

Step-1: The collection VA of intervals including the empty set is *partly ordered*. The ordering relation $[a,b]\leq_{VA}[c,d]$ is equivalent to "$c\leq_A a$ and $b\leq_A d$". An eligible interval notation for the empty set O_{VA} is $O_{VA}=[I_A,O_A]$.

Step-2: Any two intervals $[a,b]$, $[c,d]$ in VA have a least upper bound, in particular it is $[a,b]\vee_{VA}[c,d]=[a\wedge_A c,b\vee_A d]$.

Step-3: Any two intervals $[a,b]$, $[c,d]$ in VA have a greatest lower bound, in particular it is $[a,b]\wedge_{VA}[c,d]=[a\vee_A c,b\wedge_A d]$ if $a\vee_A c\leq_A b\wedge_A d$, otherwise it is $[a,b]\wedge_{VA}[c,d]=O_{VA}$.

Step-4: The implied lattice $<\mathsf{VA},\leq_{VA}>$ is a complete lattice, that is any subset of VA has both a least upper bound and a greatest lower bound in $<\mathsf{VA},\leq_{VA}>$.

Note that the *lattice monotone properties* "$x\leq_A y \Rightarrow x\wedge_A z\leq_A y\wedge_A z$, and $x\leq_A y \Rightarrow x\vee_A z\leq_A y\vee_A z$" [1] will be used extensively in the proofs below.

Proof of Step-1

Recall that a non-empty interval $\Delta\in\mathsf{VA}$ has been defined as the set $[a,b]\triangleq\{x\in<\mathsf{A},\leq_A> : a\leq_A x\leq_A b\}$. Therefore the collection VA of intervals including the empty set O_{VA} is *partly ordered* under the conventional set-inclusion relation, the latter relation is denoted herein by $\leq_{VA}$. Note that a set is called partly ordered when the *reflexive*, *antisymmetry*, and *transitive* laws hold [1].

For non-empty intervals $[a,b]$, $[c,d]\in\mathsf{VA}$ the partial ordering relation $[a,b]\leq_{VA}[c,d]$ is equivalent to $c\leq_A a\leq_A b\leq_A d$. Aiming at introducing the standard interval notation for the empty set we "relaxed" the aforementioned equivalence relation in VA and we have replaced it by "$[a,b]\leq_{VA}[c,d] \Leftrightarrow c\leq_A a$ and $b\leq_A d$". All the intervals in VA including the empty interval O_{VA} whose interval notation is searched for, say it is $O_{VA}=[e_1,e_2]$, are required to satisfy the previous "relaxed" equivalence relation.

On the one hand, the non-empty intervals already satisfy the "relaxed" equivalence relation. On the other hand for the empty interval, the "relaxed" equivalence relation implies $[e_1,e_2]\leq_{VA}[a,b] \Leftrightarrow a\leq_A e_1$ and $e_2\leq_A b$, for all $[a,b]\in\mathsf{VA}$. Therefore we propose the following interval notation for the empty interval $O_{VA}=[e_1,e_2]=[I_A,O_A]$.

We show now that denoting the empty interval by $O_{VA}=[I_A,O_A]$ complies with standard lattice theoretical considerations. First, we consider a non-empty interval say $\Delta=[a,b]$ where $a,b\in<\mathsf{A},\leq_A>$ and $a\leq_A b$. Because $<\mathsf{A},\leq_A>$ is a complete lattice then both the least upper bound of the set Δ, denoted by $\vee_A\Delta$, and the greatest lower bound of Δ, denoted by $\wedge_A\Delta$, exist in $<\mathsf{A},\leq_A>$. In particular it holds $\wedge_A\Delta=a$ and $\vee_A\Delta=b$. Hence a non-empty interval can be denoted by $\Delta=[\wedge_A\Delta,\vee_A\Delta]$. Second, we examine whether the equality $\Delta=[\wedge_A\Delta,\vee_A\Delta]$ holds when $\Delta=O_{VA}$. We cite part of the content of Remark 2.2 from [10] : "the greatest lower bound of the empty set in a complete lattice $<\mathsf{A},\leq_A>$ is I_A, while the least upper bound of the empty set is O_A". Hence in order to comply with Remark 2.2 in [10] the empty interval O_{VA} has to be denoted by $O_{VA}=[\wedge_A O_{VA},\vee_A O_{VA}]=[I_A,O_A]$. That is, we have reconfirmed the validity of our proposed interval notation for the empty set.

Proof of Step-2

Consider the interval $[a\wedge_A c,b\vee_A d]$. It is $a\wedge_A c\leq_A a$ and $b\leq_A b\vee_A d$, hence $[a,b]\leq_{VA}[a\wedge_A c,b\vee_A d]$. Likewise $[c,d]\leq_{VA}[a\wedge_A c,b\vee_A d]$. Therefore $[a\wedge_A c,b\vee_A d]$ is an upper bound of both $[a,b]$ and $[c,d]$. We show now that $[a\wedge_A c,b\vee_A d]$ is the least upper bound of $[a,b]$ and $[c,d]$. Towards this end assume another upper bound interval of both $[a,b]$ and $[c,d]$, say interval $[\beta,\gamma]$. The latter assumption implies both (i) $[a,b]\leq_{VA}[\beta,\gamma]\Leftrightarrow \beta\leq_A a,\ b\leq_A\gamma$, and (ii) $[c,d]\leq_{VA}[\beta,\gamma]\ \Leftrightarrow\ \beta\leq_A c,\ d\leq_A\gamma$. Employing the *lattice monotone properties*, we can conclude both $\beta\leq_A a\wedge_A c$ and $b\vee_A d\leq_A\gamma$, hence $[a\wedge_A c,b\vee_A d]\leq_{VA}[\beta,\gamma]$. In conclusion $[a,b]\vee_{VA}[c,d]= [a\wedge_A c,b\vee_A d]$.

Proof of Step-3

To calculate the $[a,b]\wedge_{VA}[c,d]$ we will need to show at first the following equivalence : $[a,b]\cap[c,d]\neq O_{VA} \Leftrightarrow a\vee_A c\leq_A b\wedge_A d$, where $\cap$ is the set-intersection operator and O_{VA} denotes the empty set. In the one direction of the previous equivalence, assume that the sets $[a,b]$ and $[c,d]$ intersect each other. Then $\exists\gamma\in$ <A,$\leq_A$>: $a\leq_A\gamma\leq_A b$ and $c\leq_A\gamma\leq_A d$. Employing the *lattice monotone properties* we conclude both $a\vee_A c\leq_A\gamma$ and $\gamma\leq_A b\wedge_A d$. Therefore $a\vee_A c\leq_A b\wedge_A d$. In the other direction of the previous equivalence, assume that $a\vee_A c\leq_A b\wedge_A d$. Then $\exists\gamma\in$ <A,$\leq_A$>: $a\leq_A a\vee_A c\leq_A\gamma\leq_A b\wedge_A d\leq_A b\ \Rightarrow\ \gamma\in[a,b]$, and $c\leq_A a\vee_A c\leq_A\gamma\leq_A b\wedge_A d\leq_A d.\ \Rightarrow\ \gamma\in[c,d]$. Therefore $[a,b]\cap[c,d]\neq O_{VA}$..

We resume now the calculation of $[a,b]\wedge_{VA}[c,d]$. On the one hand, if $a\vee_A c\not\leq_A b\wedge_A d$ then by employing the above equivalence we conclude $[a,b]\cap[c,d]=O_{VA}$., hence we infer $[a,b]\wedge_{VA}[c,d]=O_{VA}$. On the other hand, if $a\vee_A c\leq_A b\wedge_A d$ then consider the interval $[a\vee_A c,b\wedge_A d]$. It is both $a\leq_A a\vee_A c$ and $b\wedge_A d\leq_A b$, hence $[a\vee_A c,b\wedge_A d]\leq[a,b]$. Likewise $[a\vee_A c,b\wedge_A d]\leq[c,d]$. Therefore $[a\vee_A c,b\wedge_A d]$ is a lower bound of both $[a,b]$ and $[c,d]$. We show now that $[a\vee_A c,b\wedge_A d]$ is the greatest lower bound of $[a,b]$ and $[c,d]$. Towards this end assume another lower bound interval of both $[a,b]$ and $[c,d]$, say interval $[\beta,\gamma]$. The latter assumption implies both (i) $[\beta,\gamma]\leq[a,b]\ \Leftrightarrow\ a\leq\beta,\ \gamma\leq b$, and (ii) $[\beta,\gamma]\leq[c,d]\ \Leftrightarrow\ c\leq\beta,\ \gamma\leq d$. Employing the *lattice monotone properties* we conclude both $a\vee_A c\leq_A\beta$ and $\gamma\leq_A b\wedge_A d$, therefore we infer $[\beta,\gamma]\leq_{VA}[a\vee_A c,b\wedge_A d]$. In conclusion $[a,b]\wedge_{VA}[c,d]= [a\vee_A c,b\wedge_A d]$.

Proof of Step-4

Considering jointly the results of Step-1, Step-2, Step-3 and applying the definition of a lattice given in the beginning of section 2 we conclude that <VA,$\leq_{VA}$> is a lattice.

Let $\{[a_i,b_i]\}_{i\in I}$ denote a collection of intervals in <VA,$\leq_{VA}$>, where I is an index set. We show now that $\vee_{VA}\{[a_i,b_i]\}_{i\in I}$ exists in <VA,$\leq_{VA}$> in particular it is $\vee_{VA}\{[a_i,b_i]\}_{i\in I}= [\wedge_A\{a_i\}_{i\in I},\ \vee_A\{b_i\}_{i\in I}]$, where both $\wedge_A\{a_i\}_{i\in I}$ and $\vee_A\{b_i\}_{i\in I}$ exist in the complete lattice <A,$\leq_A$>. We show first that $[\wedge_A\{a_i\}_{i\in I},\ \vee_A\{b_i\}_{i\in I}]$ is an upper bound of all $[a_i,b_i]$, $i\in I$, and then we show that $[\wedge_A\{a_i\}_{i\in I},\ \vee_A\{b_i\}_{i\in I}]$ is their least upper bound. The details of the proof the same as in the Proof of Step-2, and they will not be repeated. In conclusion $\vee_{VA}\{[a_i,b_i]\}_{i\in I}= [\wedge_A\{a_i\}_{i\in I},\ \vee_A\{b_i\}_{i\in I}]$.

Likewise, by following the same line of arguments as in the Proof of Step-3 we can show that $\wedge_{VA}\{[a_i,b_i]\}_{i\in I}$ exists in <VA,$\leq_{VA}$>. In particular, $\wedge_{VA}\{[a_i,b_i]\}_{i\in I}= [\vee_A\{a_i\}_{i\in I},\ \wedge_A\{b_i\}_{i\in I}]$ if $\vee_A\{a_i\}_{i\in I}\leq\wedge_A\{b_i\}_{i\in I}$, otherwise $\wedge_{VA}\{[a_i,b_i]\}_{i\in I}=O_{VA}$..

In conclusion we infer that <VA,$\leq_{VA}$> is a complete lattice.

Proof of Proposition 12

Let x,y be in <A×A,$\leq_{A\times A,1}$> with $x=(a,b)$, and $y=(c,d)$, where $a,b,c,d\in$ <A,$\leq_A$>, and let v be a valuation on <A,$\leq_A$>. Then

$V(x)+V(y)=V(a,b)+V(c,d)=[v(a)+v(b)]+[v(c)+v(d)]=[v(a)+v(c)]+[v(b)+v(d)]=[v(a\vee_{\mathsf{A}}c)+v(a\wedge_{\mathsf{A}}c)]+[v(b\vee_{\mathsf{A}} d)+v(b\wedge_{\mathsf{A}}d)]=[v(a\vee_{\mathsf{A}}c)+v(b\vee_{\mathsf{A}}d)]+[v(a\wedge_{\mathsf{A}}c)+v(b\wedge_{\mathsf{A}}d)]=V(a\vee_{\mathsf{A}}c,b\vee_{\mathsf{A}}d)+V(a\wedge_{\mathsf{A}}c,b\wedge_{\mathsf{A}}d)=V((a,b)\vee_{\mathsf{A}\times\mathsf{A},1}(c,d))+ V((a,b)\wedge_{\mathsf{A}\times\mathsf{A},1}(c,d))= V(x\vee_{\mathsf{A}\times\mathsf{A},1}y)+V(x\wedge_{\mathsf{A}\times\mathsf{A},1}y)$, that is V is a valuation on the lattice $<\mathsf{A}\times\mathsf{A},\leq_{\mathsf{A}\times\mathsf{A},1}>$.

The partial ordering relation in lattice $<\mathsf{A}\times\mathsf{A},\leq_{\mathsf{A}\times\mathsf{A},1}>$, that is $x\leq_{\mathsf{A}\times\mathsf{A},1}y \Leftrightarrow (a,b)\leq_{\mathsf{A}\times\mathsf{A},1}(c,d)$, implies $a\leq_{\mathsf{A}}c$ and $b\leq_{\mathsf{A}}d$. Therefore if v is, in addition, a monotone valuation in $<\mathsf{A},\leq_{\mathsf{A}}>$ then $x\leq_{\mathsf{A}\times\mathsf{A},1}y$ Þ $a\leq_{\mathsf{A}}c$ and $b\leq_{\mathsf{A}}d \Rightarrow v(a)\leq v(c)$, $v(b)\leq v(d)$, that is $v(a)+v(b)\leq v(c)+v(d) \Rightarrow V(x)\leq V(y)$, hence V is a monotone valuation on $<\mathsf{A}\times\mathsf{A},\leq_{\mathsf{A}\times\mathsf{A},1}>$. Moreover and likewise, a positive valuation v on $<\mathsf{A},\leq_{\mathsf{A}}>$ implies the positive valuation $V(a,b)=v(a)+v(b)$ on $<\mathsf{A}\times\mathsf{A},\leq_{\mathsf{A}\times\mathsf{A},1}>$.

REFERENCES

[1] G. Birkhoff, *Lattice Theory*, American Mathematical Society, Colloquium Publications, vol. 25, 1967.

[2] A. Blumer, A. Ehrenfeucht, D. Haussler, and M.K. Warmuth, "Learnability and the Vapnik-Chervonenkis dimension", *Journal of the ACM*, vol. 36 (4), pp. 929-965, 1989.

[3] G. Carpenter, and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine", *Computer Vision, Graphics and Image Understanding*, vol. 37, pp. 54-115, 1987.

[4] G. Carpenter, and S. Grossberg, "ART2: Stable Self-Organization of Pattern Recognition Codes for Analog Input Patterns", *Applied Optics*, vol. 26, pp. 4919-4930, 1987.

[5] G. Carpenter, and S. Grossberg, "ART3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures", *Neural Networks*, vol. 3, pp. 129-152, 1990.

[6] G. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D. Rosen, "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps", *IEEE Trans. on Neural Networks*, vol. 3, no. 5, pp. 698-713, 1992.

[7] G. Carpenter, S. Grossberg, and J.H. Reynolds, "ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network", *Neural Networks*, vol. 4, pp. 565-588, 1991.

[8] G. Carpenter, S. Grossberg, and J.H. Reynolds, "A Fuzzy ARTMAP Nonparametric Probability Estimator for Nonstationary Pattern Recognition Problems", *IEEE Trans. on Neural Networks*, vol. 6, no. 6, pp. 1330-1336, 1995.

[9] G. Carpenter, S. Grossberg, and D. Rosen, "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system", *Neural Networks*, vol. 4, pp. 759-771, 1991.

[10] B.A. Davey, and H.A. Priestley, *Introduction to Lattices and Order*, Cambridge University Press, 1990.

[11] M. Delgado, A.F. Gomez-Skarmeta, and F. Martin, "A Fuzzy Clustering-Based Rapid Prototyping for Fuzzy Rule-Based Modeling", *IEEE Trans. on Fuzzy Systems*, vol. 5, no. 2, pp 223-233, 1997.

[12] E.A. Edmonds, "Lattice Fuzzy Logics", *International Journal of Man-Machine Studies*, vol. 13 (4), pp. 455-465, 1980.

[13] S.E. Fahlman, and M. White, Carnegie Mellon University (CMU) Repository of Neural Net Benchmarks, [http://www.boltz.cs.cmu.edu/].

[14] B. Fritzke, "Growing Cell Structures - A Self-Organizing Network for Unsupervised and Supervised Learning", *Neural Networks*, vol. 7, no. 9, pp. 1441-1460, 1994.

[15] B.R. Gaines, "Fuzzy and Probability Uncertainty Logics", *Information and Control*, vol. 38, pp. 154-169, 1978.

[16] J.A. Goguen , "L-Fuzzy Sets", *J. Mathematical Analysis and Applications*, vol. 18, pp. 145-174, 1967.

[17] M. Georgiopoulos, H. Fernlund, G. Bebis, and G.L. Heileman, "Order of Search in Fuzzy ART and Fuzzy ARTMAP: Effect of the Choice Parameter", *Neural Networks*, vol. 9, no. 9, pp. 1541-1559, 1996.
[18] P.R. Halmos, *Naive Set Theory*, Van Nostrand, New York, 1960.
[19] J. Huang, M. Georgiopoulos, and G.L. Heileman, "Fuzzy ART Properties", *Neural Networks*, vol. 8, no. 2, pp. 203-213, 1995.
[20] H. Ishibuchi, R. Fujioka, and H. Tanaka, "Neural Networks That Learn from Fuzzy If-Then Rules", *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 2, pp. 85-97, 1993.
[21] J.-S. R. Jang, and C.-T. Sun, "Neuro-Fuzzy Modeling and Control", *Proceedings of the IEEE*, vol. 83, no. 3, pp. 378-406, 1995.
[22] V.G. Kaburlasos, "Adaptive Resonance Theory with Supervised Learning and Large Database Applications", Dept Electrical Engineering, University of Nevada Reno, Ph.D. Dissertation, 1992.
[23] V.G. Kaburlasos, and V. Petridis, "Fuzzy Lattice Neurocomputing (FLN) : A Novel Connectionist Scheme for Versatile Learning and Decision Making by Clustering", *International Journal of Computers and Their Applications*, vol. 4, no. 2, pp. 31-43, 1997.
[24] C.F. Juang, and C.T. Lin, "An On-Line Self-Constructing Neural Fuzzy Inference Network and Its Applications", *IEEE Trans. on Fuzzy Systems*, vol. 6, no. 1, pp. 12-32, 1998.
[25] G.J. Klir, and T.A Folger, *Fuzzy Sets, Uncertainty, and Information*, Prentice-Hall International Editions, 1988.
[26] C.J. Lin, and C.T. Lin, "An ART-Based Fuzzy Adaptive Learning Control Network", *IEEE Trans. on Fuzzy Systems*, vol. 5, no. 4, pp. 477-496, 1997.
[27] Z.Q. Liu, and F. Yan, "Fuzzy Neural Network in Case-Based Diagnostic System", *IEEE Trans. on Fuzzy Systems*, vol. 5, no. 2, pp. 209-222, 1997.
[28] B.C. Lovell, and A.P. Bradley, "The Multiscale Classifier", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 124-137, 1996.
[29] C.J. Merz, and P.M. Murphy, UCI Repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science, 1996. [http://www.ics.uci.edu/ ˜mlearn/MLRepository.html].
[30] S. Mitra, R.K. De, and S.K. Pal, "Knowledge-Based Fuzzy MLP for Classification and Rule Generation", *IEEE Trans. on Neural Networks*, vol. 8, no. 6, pp. 1338-1350, 1997.
[31] S. Mitra, and S.K. Pal, "Fuzzy Multi-Layer Perceptron, Inferencing and Rule Generation", *IEEE Trans. on Neural Networks*, vol. 6, no. 1, pp. 51-63, 1995.
[32] N. Moray, "Intelligent aids, mental models, and the theory of machines", *International Journal of Man-Machine Studies*, vol. 27, pp. 619-629, 1987.
[33] A.V. Oppenheim, and J.S. Lim, "The Importance of Phase in Signals", *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529-541, 1981.
[34] S.K. Pal, and S. Mitra, "Multilayer Perceptron, Fuzzy Sets, and Classification", *IEEE Trans. on Neural Networks*, vol. 3, no. 5, pp. 683-697, 1992.
[35] T.J. Parker, A.J. Harrison, P.N. Brett, and T. Thomas, "Towards a Mechatronic Tool for the Epidural Procedure", *Mechatronics and Machine Vision in Practice (M2ViP95)*, Hong-Kong, 1995.
[36] V. Petridis V, and V.G. Kaburlasos, "FLN: A Fuzzy Lattice Neurocomputing Scheme for Clustering", *World Congress on Neural Networks '96*, San Diego CA, pp. 942-945, 1996.
[37] V. Petridis, V.G. Kaburlasos, P. Brett, T. Parker, and J.C.C. Day, "Two Level Fuzzy Lattice (2L-FL) Supervised Clustering : A New Method for Soft Tissue Identification in Surgery", *CESA'96 IMACS Multiconference*, Lille France, pp. 232-237, 1996.
[38] D.E. Rutherford, *Introduction to Lattice Theory*, Oliver and Boyd Ltd., Edinburgh, Great Britain, 1965.
[39] P.K. Simpson, "Fuzzy Min-Max Neural Networks - Part1: Classification", *IEEE Trans. on Neural Networks*, vol. 3, no. 5, pp. 776-786, 1992.
[40] P.K. Simpson, "Fuzzy Min-Max Neural Networks - Part2: Clustering", *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 1, pp. 32-45, 1993.

[41] B.P. Straughan, *Spectroscopy*, B.P. Straughan & S. Walker eds, vol. 2, Chapman & Hall, London, pp. 138-265, 1976.
[42] T. Tagaki, and M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control", *IEEE Trans. on Syst., Man, and Cybern.*, vol. 15, no. 1, pp.116-132, 1985.
[43] A-H. Tan, "Adaptive Resonance Associative Map", *Neural Networks*, vol. 8, no. 3, pp. 437-446, 1995.
[44] A-H. Tan, "Cascade ARTMAP: Integrating Neural Computation and Symbolic Knowledge Processing", *IEEE Trans. on Neural Networks*, vol. 8, no. 2, pp. 237-250, 1997.
[45] L.G. Valiant, "A theory of the learnable", *Communications of the ACM*, vol. 27 (11), pp. 1134-1142, 1984.
[46] S. Waugh, "Extending and benchmarking Cascade-Correlation", Dept of Computer Science, University of Tasmania, Ph.D. Dissertation, 1995.
[47] J.R. Williamson, "Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps", *Neural Networks*, vol. 9 (5), pp. 881-897, 1996.
[48] L.A. Zadeh, "Fuzzy Sets", *Information and Control*, vol. 8, pp. 338-353, 1965.
[49] L.A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, vol. 1, pp. 3-28, 1978.
[50] J. Zimmermann, *Fuzzy Set Theory and Its Applications*, 2nd edition, Kluwer Academic Publishers, 1991.