A Granular Extension of the Fuzzy-ARTMAP (FAM) Neural Classifier Based on Fuzzy Lattice Reasoning (FLR)

Vassilis G. Kaburlasos* and S.E. Papadakis

Technological Educational Institution of Kavala
Department of Industrial Informatics
GR-65404 Kavala, Greece
Emails: {vgkabs,spap}@teikav.edu.gr

Corresponding Author

Name: Vassilis G. Kaburlasos

Postal Address: Department of Industrial Informatics
Technological Educational Institution of Kavala
GR 654 04 Kavala
Greece

Phone Number: +30 (2510) 462-320

Fax Number: +30 (2510) 462-348

E-mail Address: vgkabs@teikav.edu.gr

* Corresponding Author Fax. no. + 30 (2510) 462-348, E-mail address: vgkabs@teikav.edu.gr

A Granular Extension of the Fuzzy-ARTMAP (FAM) Neural Classifier Based on Fuzzy Lattice Reasoning (FLR)

Abstract

The fuzzy lattice reasoning (FLR) classifier was introduced lately as an advantageous enhancement of the fuzzy-ARTMAP (FAM) neural classifier in the Euclidean space $\mathbb{R}^N$. This work extends FLR to space $\mathbb{F}^N$, where $\mathbb{F}$ is the granular data domain of Fuzzy Interval Numbers (FINs) including (fuzzy) numbers, intervals, and cumulative distribution functions. Based on a fundamentally improved mathematical notation this work proposes novel techniques for dealing, rigorously, with imprecision in practice. We demonstrate a favorable comparison of our proposed techniques with alternative techniques from the literature in an industrial prediction application involving digital images represented by histograms. Additional advantages of our techniques include a capacity to represent statistics of all orders by a FIN, an introduction of tunable (sigmoid) nonlinearities, a capacity for effective data processing without any data normalization, an induction of descriptive decision-making knowledge (rules) from the training data, and the potential for input variable selection.

Keywords: Digital image histogram; Fuzzy-ARTMAP (FAM); Fuzzy lattice reasoning (FLR); Granular computing; Industrial classification application

1. Introduction

It turns out that disparate types of data including logic values, numbers, sets, symbols, and graphs are partially(lattice)-ordered. In conclusion, a cross-fertilization was proposed, lately, in Computational Intelligence (CI) towards unified knowledge representation and modelling based on lattice theory with emphasis on clustering, classification, and regression [19].

Especially successful, within CI, was the employment of lattice theory in neural computing based on either the lattice (partial) ordering relation [19], [26], [27] or the lattice algebraic operations *meet* and *join* [38], [39], [40], [47]. In the interest of simplicity, the former employment of lattice theory is called here "order-based" (employment), whereas the latter one is called "algebra-based". On one hand, an algebra-based employment of lattice theory in neural computing typically involves morphological neural networks, which grew out of the theory of image algebra [37], [41], [44], [45]; the latter neural networks were extended, lately, based on fuzzy set theory techniques [46], [48]. On the other hand, an order-based

employment of lattice theory in neural computing has stemmed from the *adaptive resonance theory*, or *ART* for short, as explained next.

ART has evolved from studies of the brain and mind during the 1970s [16], [17]. In conclusion, a biologically inspired neural paradigm was proposed for clustering binary patterns [6]. An analog (pattern) extension of ART, namely *fuzzy-ART*, was proposed later in the unit hypercube [7]. The corresponding neural network for classification is *fuzzy-ARTMAP*, or *FAM* for short [8]. It turns out that FAM operates by conditionally inducing lattice-ordered hyperboxes from point inputs in the N-dimensional unit hypercube [19].

The original work on ART has spurred a lasting research activity. A number of authors have successfully employed FAM in various data-mining and information-processing applications of neural computing [1], [9], [10], [51], [57]. In the meanwhile, the min-max neural networks were introduced for clustering and classification based on hyperboxes [42], [43]. Note that, lately, "granular computing" [2] was pursued by min-max computing [3]. We point out that neither FAM nor min-max neural networks employ lattice theory explicitly. Nevertheless, different authors have pursued min-max computing based on mathematical morphology applications of lattice theory for pattern recognition [34], [55].

Yet, different authors have introduced *fuzzy lattice neurocomputing* (FLN) models applicable in disparate partially(lattice)-ordered data domains. In particular, the σ-FLNMAP neural network was introduced as an extension of FAM to a lattice data domain [26]. The corresponding learning/generalization algorithm of σ-FLNMAP, namely *fuzzy lattice reasoning* (FLR) algorithm, was detailed with emphasis to rule induction [21]. In addition, the FLR has enhanced σ-FLNMAP in two ways. First, by applicability to the whole Euclidean space $\mathbb{R}^N$ and, second, by an introduction of tunable (sigmoid) nonlinearities.

Of particular interest in practice is the totally-ordered lattice $(\mathbb{R},\leq)$ of real numbers, which has emerged "historically" from the conventional measurement process of successive comparisons [19]. It is known that $(\mathbb{R},\leq)$ gives rise to a hierarchy of lattices including, ultimately, the lattice $(\mathbb{F},\leq)$ of *fuzzy interval numbers*, or FINs for short [20], [31]. Note that a FIN is a unifying data representation including (fuzzy) numbers, intervals, and cumulative distribution functions [18], [19], [20], [22], [24], [25], [31]; therefore, a FIN can be interpreted as an information *granule* towards *granular computing* as explained next.

Granular computing is an emerging computational paradigm [2], which is roughly defined as a set of theories, methodologies, techniques and tools that make use of *information granules*. The latter can be conceived as a collection of entities grouped together by similarity, functional adjacency, coherency, etc. The basic notions and principles of granular computing have appeared under different names in related fields such as *information hiding* in

programming, *granularity* in artificial intelligence, *divide-and-conquer* in theoretical computer science, interval computing, cluster analysis, fuzzy and rough set theories, etc. [3], [28], [33]. This paper pursues *granular computing* rigorously based on FINs. In particular, this paper builds on previous work including substantial novelties as explained next.

A previous version of FLR dealt only with N-dimensional hyperboxes [21]; whereas, this work extends the applicability of FLR to the space of FINs. Note that recent works have employed FINs towards classification [25], [31]. Nevertheless, the latter employ *interval-FINs* characterized by both theoretical and practical drawbacks regarding FINs with largest height less than 1. More specifcally, on one hand, the work in [25] uses an agglomerative clustering algorithm based on a *metric* function, whereas the work here employs a FLR algorithm based on an *inclusion measure* function. On the other hand, the work in [31] employs a FLR algorithm based on a *similarity measure* function, whereas the FLR here employs an *inclusion measure* function for carrying out tunable inferences.

Another, substantial novelty of this work includes a novel mathematical formulation regarding FINs. More specifically, both this as well as previous works represent a FIN as a collection of *generalized intervals*. Nevertheless, previous works represented a generalized interval as a positive/negative "pulse-shaped" function of height $h \in (0,1]$ resulting in FINs with both positive and negative membership functions [18], [19], [22], [23], [24], [25], [31]. This work abandons negative (FIN) membership functions altogether in order to avoid confusion. Hence, a FIN is represented here as a collection of only positive generalized intervals, which may be interpreted as conventional *α-cuts* in fuzzy set theory. Furthermore, fuzzy set theory's "resolution identity theorem" [56] is invoked here as a valid mathematical result in order to represent a general FIN by either a set of (positive) generalized intervals or a membership function [20]. Another advantage of our proposed (novel) mathematical formulation is that it permits a unified treatment of the Euclidean space $\mathbf{R}^N$ with its isomorphic unit hypercube as explained below. In addition, no data normalization in the interval [0,1] is required here due to the employment of sigmoid (positive valuation) functions. Moreover, a sigmoid positive valuation here suggests a method for input variable selection. Furthermore, the FLR algorithm here can deal "in principle" with both "missing" data and "don't care" data.

Previous work has used a *mass* function in order to define a positive valuation function by the computation of an integral [22], [23], [24]. Despite an insightful interpretation of a mass function as a weight function, the calculation of an integral is computationally expensive. Therefore, in the interest of simplicity, here we employ directly a parametric (sigmoid) *strictly increasing* function as a (tunable) positive valuation function.

Furthermore, we demonstrate new experimental results regarding an industrial prediction application, where a FIN represents a digital image histogram. The results compare favorably with the results by different k-nearest-neighbor (kNN) classifiers. Moreover, the FLR classifier induces descriptive, decision-making knowledge (rules) for classification.

The layout is as follows. Section 2 presents a hierarchy of popular lattices including a novel mathematical formulation, which is employed in section 3 for introducing a granular extension of the FAM classifier. Section 4 describes comparatively an industrial application. Section 5 concludes by summarizing the contribution of this work. Finally, the Appendix summarizes basic lattice theory notions and useful results including a novel theorem proof.

2. A hierarchy of popular lattices

Basic lattice theory notions and useful results, including a novel theorem proof, are shown in the Appendix in order to make this work "self-contained". This section focuses on the totally-ordered lattice $(\mathbb{R},\leq)$ of real numbers as well as on a hierarchy of popular lattices stemming from $(\mathbb{R},\leq)$ thus introducing a fundamentally improved mathematical notation.

2.1. The complete lattice $(\Delta,\leq)$ of generalized intervals and extensions

According to some authors, lattice $(\mathbb{R},\leq)$ is noncomplete [12]. For technical reasons, pertaining to compatibility with previous work, we augmented $(\mathbb{R},\leq)$ to a complete lattice with *least* and *greatest* elements denoted, respectively, by $O=-\infty$ and $I=+\infty$. Using a strictly increasing function f: $\mathbb{R}\rightarrow[0,1]$, e.g. a sigmoid $f(x)= 1/(1+e^{-x})$, we can establish an (order) isomorphism $(\mathbb{R},\leq) \cong (\mathbb{I}=[0,1],\leq)$ – For definition of "isomorphism" see in the Appendix.

Consider the complete product lattice $(\Delta,\leq) = (\mathbb{R},\leq^{\partial})\times(\mathbb{R},\leq) = (\mathbb{R}\times\mathbb{R},\geq\times\leq)$ of *generalized intervals*. A *generalized interval* will be denoted by $[x,y]$, $x,y\in\mathbb{R}$. The corresponding *meet* and *join* in lattice $(\Delta,\leq)$ are given, respectively, by $[a,b]\wedge[c,d] = [a\vee c,b\wedge d]$ and $[a,b]\vee[c,d] = [a\wedge c,b\vee d]$ – We point out that $a\wedge c$ denotes the *minimum* of real numbers a and c, whereas $a\vee c$ denotes the corresponding *maximum*. The set of *positive* (*negative*) generalized intervals $[a,b]$, characterized by $a\leq b$ ($a>b$), will be denoted by Δ_+ (Δ_-). Note that lattice $(\Delta_+,\leq)$ of *positive* generalized intervals is isomorphic to the lattice $(\tau(\mathbb{R}),\leq)$ of intervals (sets) in the set $\mathbb{R}$, i.e. $(\tau(\mathbb{R}),\leq) \cong (\Delta_+,\leq)$. We have augmented lattice $(\tau(\mathbb{R}),\leq)$ by a *least* (empty) interval, denoted by $O=[+\infty,-\infty]$, as explained in [26] – Note that a *greatest* interval $I=[-\infty,+\infty]$ already exists in $\tau(\mathbb{R})$. Hence, the complete lattice $(\tau_O(\mathbb{R})= \tau(\mathbb{R})\cup\{O\},\leq)$ emerged. The previous analysis also applies to isomorphic lattice $(\mathbb{I},\leq) \cong (\mathbb{R},\leq)$ resulting in the complete lattice $(\tau_O(\mathbb{I})=\tau(\mathbb{I})\cup\{O\},\leq) \cong (\tau_O(\mathbb{R}),\leq)$. In the interest of simplicity we use identical symbols O

and I to denote the least and greatest element, respectively, in all complete lattices including $(\mathsf{R},\leq) \cong (\mathsf{I},\leq)$ as well as $(\tau_O(\mathsf{R}),\leq) \cong (\tau_O(\mathsf{I}),\leq)$.

A (*strictly*) *decreasing bijective*, that latter means "one-to-one", function $\theta_{\mathsf{R}}: \mathsf{R}\to\mathsf{R}$ implies an isomorphism $(\mathsf{R},\leq) \cong (\mathsf{R},\geq)$; i.e. $x<y \Leftrightarrow \theta_{\mathsf{R}}(x)>\theta_{\mathsf{R}}(y)$, for $x,y\in\mathsf{R}$. Furthermore, a *strictly increasing* function $v_{\mathsf{R}}: \mathsf{R}\to\mathsf{R}$ is a positive valuation in lattice $(\mathsf{R},\leq)$. Therefore, function $v_{\Delta}: \Delta\to\mathsf{R}$ given by $v_{\Delta}([a,b])= v_{\mathsf{R}}(\theta_{\mathsf{R}}(a))+v_{\mathsf{R}}(b)$ is a positive valuation in lattice $(\Delta,\leq)$ [21]. It follows metric $d_{\Delta}: \Delta\times\Delta\to\mathsf{R}_0^+$ given by $d_{\Delta}([a,b],[c,d]) = [v_{\mathsf{R}}(\theta_{\mathsf{R}}(a\wedge c))-v_{\mathsf{R}}(\theta_{\mathsf{R}}(a\vee c))] + [v_{\mathsf{R}}(b\vee d)-v_{\mathsf{R}}(b\wedge d)]$. Note that metric d_{Δ} is valid, in particular, in lattice $(\Delta_+\cup\{O\},\leq) \cong (\tau_O(\mathsf{R}),\leq) \cong (\tau_O(\mathsf{I}),\leq)$. For reasons explained in the Appendix (see Theorem A.10) our interest here is, in particular, in positive valuation functions $v_{\mathsf{R}}: \mathsf{R}\to\mathsf{R}_0^+$ such that $0= v_{\mathsf{R}}(O=-\infty) < v_{\mathsf{R}}(I=+\infty) <+\infty$.

It is straightforward to define an inclusion measure in the complete lattice $(\tau_O(\mathsf{R}),\leq)$ as follows. Consider the aforementioned positive valuation function $v_{\Delta}: \Delta\to\mathsf{R}_0^+$ given by $v_{\Delta}([a,b])= v_{\mathsf{R}}(\theta_{\mathsf{R}}(a))+v_{\mathsf{R}}(b)$ in complete lattice $(\Delta,\leq)$. Since the conditions of Theorem A.10 (in the Appendix) are satisfied, there follow two inclusion measures, namely $k(.,.)$ and $s(.,.)$. The latter inclusion measures are valid, in particular, in lattice $(\Delta_+\cup\{O\},\leq) \cong (\tau_O(\mathsf{R}),\leq) \cong (\tau_O(\mathsf{I}),\leq)$.

Functions $\theta_{\mathsf{R}}(.)$ and $v_{\mathsf{R}}(.)$ can be selected in many different ways. For instance, choosing both $\theta_{\mathsf{R}}(x)= -x$ and $v_{\mathsf{R}}(.)$ such that $v_{\mathsf{R}}(x) = -v_{\mathsf{R}}(-x)$ it follows positive valuation function $v_{\Delta}([a,b]) = v_{\mathsf{R}}(b) - v_{\mathsf{R}}(a)$; hence, it follows metric $d_{\Delta}([a,b],[c,d])= [v_{\mathsf{R}}(a\vee c)-v_{\mathsf{R}}(a\wedge c)] + [v_{\mathsf{R}}(b\vee d)-v_{\mathsf{R}}(b\wedge d)]$ [22]. In particular, for $\theta_{\mathsf{R}}(x)= -x$ and $v_{\mathsf{R}}(x)= x$ it follows metric $d_{\Delta}([a,b])= |a-c| + |b-d|$. In general, *parametric* functions $\theta_{\mathsf{R}}(.)$ and $v_{\mathsf{R}}(.)$ imply tunable nonlinearities, where the corresponding parameters may be estimated optimally by various techniques such as stochastic search including genetic algorithms.

Another lattice of practical interest is the Cartesian product $(\tau_O(\mathsf{R}^N)= [\tau(\mathsf{R})]^N\cup\{O\},\leq)$. Both a metric and an inclusion measure function can be defined in the complete lattice $(\tau_O(\mathsf{R}^N),\leq)$, of N-dimensional hyperboxes, based on the Cartesian product $(\Delta,\leq)\times\ldots\times(\Delta,\leq) = (\Delta,\leq)^N$, in two different ways. First, one can regard $(\Delta,\leq)^N$ as a (single) complete lattice $(\Delta^N,\leq)$ with positive valuation function $v_{\Delta^N}: \Delta^N\to\mathsf{R}_0^+$ given by $v_{\Delta^N}([a_1,b_1]\times\ldots\times[a_N,b_N]) = v_{\Delta,1}([a_1,b_1])+\ldots+v_{\Delta,1}([a_N,b_N]) = \sum_{i=1}^{N}[v_{\mathsf{R},i}(\theta_{\mathsf{R},i}(a_i))+v_{\mathsf{R},i}(b_i)]$; hence, a metric and an inclusion measure function can be defined, respectively, by equation (E1) and Theorem A.10 in the Appendix. Second, one can regard $(\Delta,\leq)^N$ as a product of N *constituent* lattices. Hence, a metric and an inclusion measure can be defined, respectively, by equation (E2) and Theorem A.11 in the Appendix.

2.2. The lattice $(\mathsf{F},\leq)$ of fuzzy interval numbers (FINs) and extensions

A fundamental result of *fuzzy set theory* is the "resolution identity theorem", which states that a fuzzy set can, equivalently, be represented either by its membership function or by its α-cuts [56]. The aforementioned theorem was given little attention in practice, to-date. However, some authors have capitalized on it by designing effective as well as efficient fuzzy inference systems (FIS) based on α-cuts of fuzzy numbers, i.e. based on intervals in $\tau(\mathsf{R})$ [49], [50]; more specifically, advantages include faster (parallel) data processing "level-by-level" as well as "orders-of-magnitude smaller" computer memory requirements for representing, equivalently, fuzzy sets with arbitrary membership functions. This work builds on the resolution identity theorem as follows.

In the first place, we drop a possibilistic interpretation for a (fuzzy) membership function regarding fuzzy numbers. Then, we consider the corresponding "α-cuts representation". In conclusion, a *fuzzy interval number* (*FIN*) emerges as a function F: $(0,1]\rightarrow(\Delta_{+}\cup\{O\},\leq) \cong (\tau_{O}(\mathsf{R}),\leq)$ as detailed below. A more general number type is defined first.

Definition 1. A *generalized interval number* (*GIN*) is a function f: $(0,1]\rightarrow\Delta$. ■

Let G denote the set of GINs. It follows that $(\mathsf{G},\leq)$ is a complete lattice since $(\mathsf{G},\leq)$ is the Cartesian product of complete lattices $(\Delta,\leq)$. Our interest here focuses on the *sublattice* of *fuzzy interval numbers* defined next – For definition of "sublattice" see in the Appendix.

Definition 2. A *fuzzy interval number* (*FIN*) F is a GIN such that both $F(h)\in[\Delta_{+}\cup\{O\}]$ and $h_1\leq h_2 \Rightarrow F(h_1)\geq F(h_2)$, for all $h\in(0,1]$. ■

For graphical illustrations regarding FIN interpretations the interested reader may refer to [18], [24], [35]. Let F denote the set of FINs. Apparently, it is $\mathsf{F}\subseteq\mathsf{G}$. Moreover, $F,E\in\mathsf{F}$ imply both $(F\vee E)\in\mathsf{F}$ and $(F\wedge E)\in\mathsf{F}$ as follows: $h_1\leq h_2$ implies both $F(h_1)\geq F(h_2)$ and $E(h_1)\geq E(h_2)$; hence, first, $(F\vee E)(h_1) = F(h_1)\vee E(h_1) \geq F(h_2)\vee E(h_1) \geq F(h_2)\vee E(h_2) = (F\vee E)(h_2) \Rightarrow (F\vee E)\in\mathsf{F}$; second, $(F\wedge E)(h_1) = F(h_1)\wedge E(h_1) \geq F(h_2)\wedge E(h_1) \geq F(h_2)\wedge E(h_2) = (F\wedge E)(h_2) \Rightarrow (F\wedge E)\in\mathsf{F}$. Therefore, poset $(\mathsf{F},\leq)$ is a lattice.

Previous work has shown that the cardinality of set F equals $\aleph_1$, that is the cardinality of the set R of real numbers [22]; in other words, there are as many FINs as there are real numbers. However, previous work [18], [19], [22], [23], [24], [25], [31] employed "box-shaped" generalized intervals of height $h\in(0,1]$. Whereas, this work employs a fundamentally improved, novel mathematical notation based on the lattice $(\Delta,\leq) = (\mathsf{R}\times\mathsf{R},\geq\times\leq)$ of generalized intervals. Hence, using the notation here, FINs can directly be associated with α-cuts of fuzzy numbers towards a wider proliferation of FINs in practical applications.

A FIN will typically be denoted by a capital letter in italics, e.g. $F \in \mathsf{F}$. Moreover, a N-tuple FIN will typically be denoted by a capital letter in bold italics, e.g. $\boldsymbol{F}=(F_1,\ldots,F_N) \in \mathsf{F}^N$.

A FIN F can be written as the set union of generalized intervals, e.g. $F= \bigcup_{h\in(0,1]} \{[a_h,b_h]\}$, where both interval-ends a_h and b_h are functions of $h \in (0,1]$. A FIN may admit different interpretations including a (fuzzy) number, an interval, and a cumulative distribution function. For instance, a FIN $F= \bigcup_{h\in(0,1]} \{[a,b]\}$ represents interval $[a,b]$ including real numbers for $a=b$. Moreover, a FIN can represent a probability distribution function such that interval $F(h)$ includes $100(1-h)\%$ of the distribution, whereas the remaining $100h\%$ is *split even* both below and above interval $F(h)$ [18], [19], [23], [24] – For general transformations between probability- and possibility- distributions the interested reader may refer to [29]. In any case, a FIN can be interpreted as an (information) *granule*. The *size* of a FIN is defined next.

Definition 3. Assuming that the following integral exists, the *size* of a FIN, with respect to a positive valuation $v_{\mathsf{R}}: \mathsf{R} \to \mathsf{R}_0^+$, is a function $Z_{\mathsf{F}}: \mathsf{F} \to \mathsf{R}_0^+$ given by $Z_{\mathsf{F}}(F) = \int_0^1 Z_{\mathsf{R}}(F(h))dh$. ■

We remark that $Z_{\mathsf{R}}(F(h))$ in Definition 3 is the size of (positive) interval $F(h)$, which (size) is computed as shown in the Appendix. Alternatively, the size of FIN $F= \bigcup_{h\in(0,1]} \{[a_h,b_h]\}$ can be computed as $Z_{\mathsf{F}}(F) = \int_0^1 d(a_h,b_h)dh$, where $d(a_h,b_h) = v(b_h)-v(a_h) = Z_{\mathsf{R}}(F(h))$ is the distance between interval $F(h) = [a_h,b_h]$ ends. The *size* $Z: \mathsf{F}^N \to \mathsf{R}_0^+$ of a N-tuple FIN $\boldsymbol{F}=(F_1,\ldots,F_N) \in \mathsf{F}^N$ is computed as $Z(\boldsymbol{F}) = Z_{\mathsf{F}}(F_1)+\ldots+ Z_{\mathsf{F}}(F_N)$.

Assuming that the following integral exists, a metric function $d_{\mathsf{F}}: \mathsf{F}\times\mathsf{F} \to \mathsf{R}_0^+$ is given by $d_{\mathsf{F}}(F_1,F_2) = \int_0^1 d_\Delta(F_1(h),F_2(h))dh$ [18], [23]; furthermore, an inclusion measure $\sigma_{\mathsf{F}}: \mathsf{F}\times\mathsf{F} \to [0,1]$ is given by $\sigma_{\mathsf{F}}(F_1,F_2) = \int_0^1 \sigma_\Delta(F_1(h),F_2(h))dh$ [18].

To enable an easy implementation of the techniques proposed in section 3, we summarize useful formulas in the following.

2.3. A list of useful formulas

It was explained at the end of section 2.1 that there are at least two different ways of defining a metric as well as an inclusion measure function in the complete lattice $(\tau_O(\mathsf{R}^N),\leq)$ of N-

dimensional hyperboxes. Hence, in this section, we define both a metric and an inclusion measure function in the complete lattice $(\mathsf{F}^N,\leq)$ of N-tuple FINs in two different ways.

First, we define a *metric* as well as an *inclusion measure* function by equation (E1) and Theorem A.10, respectively, in lattice $(\Delta^N,\leq)$.

$$\bullet\ d_1(\boldsymbol{F},\boldsymbol{E}) = \int_0^1 d_{\Delta^N}(\boldsymbol{F}(h),\boldsymbol{E}(h))dh = \int_0^1 [v_{\Delta^N}(\boldsymbol{F}(h)\vee\boldsymbol{E}(h)) - v_{\Delta^N}(\boldsymbol{F}(h)\wedge\boldsymbol{E}(h))]dh =$$

$$= \int_0^1 \sum_{i=1}^N [v_{\mathsf{R},i}(\theta_{\mathsf{R},i}(a_{i,h}\wedge c_{i,h})) - v_{\mathsf{R},i}(\theta_{\mathsf{R},i}(a_{i,h}\vee c_{i,h})) + v_{\mathsf{R},i}(b_{i,h}\vee d_{i,h}) - v_{\mathsf{R},i}(b_{i,h}\wedge d_{i,h})]dh\,,$$

$$\bullet\ \sigma_1(\boldsymbol{F},\boldsymbol{E}) = \int_0^1 \sigma_{\Delta^N}(\boldsymbol{F}(h),\boldsymbol{E}(h))dh = \int_0^1 \frac{v_{\Delta^N}(\boldsymbol{E}(h))}{v_{\Delta^N}(\boldsymbol{F}(h)\vee\boldsymbol{E}(h))}dh =$$

$$= \int_0^1 \frac{\sum_{i=1}^N [v_{\mathsf{R},i}(\theta_{\mathsf{R},i}(c_{i,h})) + v_{\mathsf{R},i}(d_{i,h})]}{\sum_{i=1}^N [v_{\mathsf{R},i}(\theta_{\mathsf{R},i}(a_{i,h}\wedge c_{i,h})) + v_{\mathsf{R},i}(b_{i,h}\vee d_{i,h})]}dh\,,\ \text{where}$$

$\boldsymbol{F}=(F_1,\ldots,F_N)=(\bigcup_{h\in(0,1]}\{[a_{1,h},b_{1,h}]\},\ldots,\bigcup_{h\in(0,1]}\{[a_{N,h},b_{N,h}]\})$, $\boldsymbol{E}=(E_1,\ldots,E_N)=(\bigcup_{h\in(0,1]}\{[c_{1,h},d_{1,h}]\},$ $\ldots,\bigcup_{h\in(0,1]}\{[c_{N,h},d_{N,h}]\})$; bijective functions $\theta_{\mathsf{R},i}$: $\mathsf{R}\to\mathsf{R}$ are strictly decreasing; and, functions $v_{\mathsf{R},i}$: $\mathsf{R}\to\mathsf{R}_0^+$ are strictly increasing with $0=v_{\mathsf{R},i}(O=-\infty)<v_{\mathsf{R},i}(I=+\infty)<+\infty$, $i\in\{1,\ldots,N\}$.

Second, we define a *metric* as well as an *inclusion measure* function by equation (E2) and Theorem A.11, respectively, in lattice $(\Delta,\leq)^N$.

$\bullet\ d_2(\boldsymbol{F},\boldsymbol{E}) = d_2((F_1,\ldots,F_N),(E_1,\ldots,E_N)) = \left[d_{\mathsf{F},1}^p(F_1,E_1)+\ldots+d_{\mathsf{F},N}^p(F_N,E_N)\right]^{1/p}$, where $p\in\mathsf{R}$, and a metric $d_{\mathsf{F},i}(E_i,F_i)$, $i=1,\ldots,N$ is computed as

$$d_{\mathsf{F}}(F,E) = \int_0^1 d_\Delta(F(h),E(h))dh = \int_0^1 d_\Delta([a_h,b_h],[c_h,d_h])dh =$$

$$= \int_0^1 [v_{\mathsf{R}}(\theta_{\mathsf{R}}(a_h\wedge c_h)) - v_{\mathsf{R}}(\theta_{\mathsf{R}}(a_h\vee c_h)) + v_{\mathsf{R}}(b_h\vee d_h) - v_{\mathsf{R}}(b_h\wedge d_h)]dh\,.$$

$\bullet\ \sigma_2(\boldsymbol{F},\boldsymbol{E}) = \sigma_2((F_1,\ldots,F_N),(E_1,\ldots,E_N)) = \lambda_1\sigma_{\mathsf{F},1}(F_1,E_1)+\ldots+\lambda_N\sigma_{\mathsf{F},N}(F_N,E_N)$, where $\lambda_1,\ldots,\lambda_N > 0$ such that $\lambda_1+\ldots+\lambda_N = 1$; moreover, an inclusion measure $\sigma_{\mathsf{F},i}(F_i,E_i)$, $i=1,\ldots,N$ is computed as

$$\sigma_F(F,E)= \int_0^1 \sigma_\Delta(F(h),E(h))dh = \int_0^1 \sigma_\Delta([a_h,b_h],[c_h,d_h])dh = \int_0^1 \frac{v_R(\theta_R(c_h))+v_R(d_h)}{v_R(\theta_R(a_h \wedge c_h))+v_R(b_h \vee d_h)}dh,$$

where bijective function θ_R: $R \to R$ is strictly decreasing, and function v_R: $R \to R_0^+$ is strictly increasing such that $0= v_R(O=-\infty) < v_R(I=+\infty) <+\infty$.

We remark that, due to the "linearity property" of the integral operator, i.e. $\int \sum [\ldots]dh = \sum \int [\ldots]dh$, it follows that $d_1(.,.)$ is a special case of $d_2(.,.)$ for p = 1. Nevertheless, inclusion measures $\sigma_1(.)$ and $\sigma_2(.)$ are, in general, different from each other for N > 1.

2.4. Practical representation issues

From a practical viewpoint we represented a FIN F in the computer memory by a L×2 matrix $[a_1\ b_1; a_2\ b_2; \ldots; a_L\ b_L]$ of real numbers, where L is a user-defined number of levels $h_1,h_2,\ldots,h_L$ such that $0 < h_1 \le h_2 \le \ldots \le h_L = 1$. In our experiments we usually used either L=16 or L=32 levels spaced equally in the interval [0,1]. Note that a similar number of 16 or 32 levels is also proposed in fuzzy inference applications based on α-level sets [49].

3. A granular extension of FAM

From an information processing point of view, the well-known FAM (fuzzy-ARTMAP) neural classifier is applicable in the *atomic lattice* of N-dimensional hyperboxes [19] – For definition of *atomic lattice* see in the Appendix. Fuzziness is introduced in FAM by calculating a (fuzzy) degree of inclusion of a trivial hyperbox, i.e. a N-dimensional point, to another hyperbox and vice-versa [26]. We point out that the hyperboxes, induced by FAM, constitute the cores of fuzzy sets in the unit hypercube. However, FAM cannot deal with fuzzy set inputs *per se*. Our work here extends the applicability of FAM to the space of FINs including (fuzzy) numbers, intervals, and cumulative distribution functions. More specifically, our work here extends the applicability of the fuzzy lattice reasoning (FLR) classifier, the latter is a lattice data domain extension of FAM [21], to the space of FINs.

3.1. Granular FLR

Granular FLR for training is presented in Fig.1 followed by *granular FLR* for testing in Fig.2. Both algorithms are applied on N-tuple FINs, namely *granules* or, equivalently, *clusters*.

The *granular FLR* is a leader-follower classifier [21], which learns rapidly in a single pass through the training data. The (granular) FLR classifier may set out learning without *a priori* knowledge; however, *a priori* knowledge can be supplied to the FLR classifier in the form of an initial number of rules in *RB* (Fig.1).

3.2. Granular FLR details

The FLR was interpreted as a rule-based classifier [21]. Regarding the *granular FLR*, in particular, a "learned" granule $\boldsymbol{E}_l \in \mathsf{F}^N$ is assigned a class label c_l, $l \in \{1,\ldots,L\}$ thus corresponding to the rule "IF granule $\boldsymbol{E}_l$ THEN class c_l", symbolically $\boldsymbol{E}_l \rightarrow c_l$.

The total number of rules is not known *a priori* but, rather, it is determined "on-line" during learning. Further training of the FLR classifier, using additional training data, does not wash away previous learning. More specifically, retraining the FLR classifier with a new data set either enhances previously learned rules (step FLR-5 in Fig. 1) or it creates new rules (step FLR-2 in Fig. 1). The maximum threshold size Z_{crit} regulates the *granularity of learning*; the latter means the number of induced rules. It turns out that, in general, larger values of Z_{crit} result in *fewer* (i.e. *more generalized*) rules, whereas smaller values of Z_{crit} result in *more* (i.e. *more specific*) rules.

Generalization can be effected in two manners. First, based on the *Assimilation Condition*, rule induction can be effected by replacing a granule $\boldsymbol{E}_J$ by a larger granule $\boldsymbol{F}_i \vee \boldsymbol{E}_J$ (Fig.1, step FLR-5). Hence, there are FINs within the larger granule $\boldsymbol{F}_i \vee \boldsymbol{E}_J$ which (FINs) are assigned category label c_J inductively, without explicit evidence. The latter is called here *Type I Generalization* and may result in granule overlapping, which can be avoided by conditionally augmenting the *Assimilation Condition* (Fig.1, step FLR-4) at the expense of longer computer processing times. Second, a pair $(\boldsymbol{E}_l, c_l)$, $l=1,\ldots,L$ defines a fuzzy set $(\mathsf{F}^N, \sigma(\boldsymbol{F} \leq \boldsymbol{E}_l))$ such that granule $\boldsymbol{E}_l$ corresponds to the *core* of fuzzy set $(\mathsf{F}^N, \sigma(\boldsymbol{F} \leq \boldsymbol{E}_l))$ – Apparently, different positive valuations imply different fuzzy membership functions; hence, generalization becomes feasible beyond core $\boldsymbol{E}_l$. The latter is called here *Type II Generalization*.

An inclusion measure $\sigma: \mathsf{F}^N \times \mathsf{F}^N \rightarrow [0,1]$ in both algorithms shown in Fig.1 and Fig.2 can be either σ_1 or σ_2 (see in section 2.3). Note that an inclusion measure σ may retain *Occam razor* semantics [21] as detailed in the following.

Let $v_{R,i}: \mathsf{R} \rightarrow \mathsf{R}_0^+$ and $\theta_{R,i}: \mathsf{R} \rightarrow \mathsf{R}$ be, respectively, a strictly increasing function and a (bijective) strictly decreasing function. A sufficient condition for Occam razor semantics is to choose $v_{R,i}$ and $\theta_{R,i}$ such that equation $v([a,b]) = C + Z([a,b])$ is satisfied, where $a,b,C \in \mathsf{R}$ with $a \leq b$, and $C>0$ constant. On one hand, regarding FAM, two popular functions $v_{R,i}$ and $\theta_{R,i}$ in the complete lattice unit-interval $\mathsf{I}=[0,1]$ are, respectively, $v_{R,i}(x) = x$ and $\theta_{R,i}(x) = 1-x$ [26]. Hence, $v_{R,i}(\theta_{R,i}(c_{i,h})) + v_{R,i}(c_{i,h}) = v_{R,i}(1-c_{i,h}) + v_{R,i}(c_{i,h}) = 1$; in conclusion, $v_{\Delta,i}([c_{i,h}, d_{i,h}]) = v_{R,i}(\theta_{R,i}(c_{i,h})) + v_{R,i}(d_{i,h}) = 1 + [v_{R,i}(d_{i,h}) - v_{R,i}(c_{i,h})] = 1 + Z_{R,i}([c_{i,h}, d_{i,h}])$. We point out that function $\theta_{R,i}(x) = 1-x$ corresponds to ART's celebrated *complement coding technique*. Apparently, choosing a different function than $\theta_{R,i}(x)=1-x$ may result in a different "coding technique"

[19], [26]. On the other hand, two popular functions $\nu_{R,i}$ and $\theta_{R,i}$ in the complete lattice R are, respectively, $\nu_{R,i}(x) = A_i/(1+e^{-\lambda_i(x-m_i)})$ and $\theta_{R,i}(x) = 2m_i - x$ [21]. Hence, $\nu_{R,i}(\theta_{R,i}(c_{i,h})) + \nu_{R,i}(c_{i,h}) = \nu_{R,i}(2m_i - c_{i,h}) + \nu_{R,i}(c_{i,h}) = A_i$; in conclusion, $\nu_{\Delta,i}([c_{i,h},d_{i,h}]) = \nu_{R,i}(\theta_{R,i}(c_{i,h})) + \nu_{R,i}(d_{i,h}) = [A_i - \nu_{R,i}(c_{i,h})] + \nu_{R,i}(d_{i,h}) = A_i + Z_{R,i}([c_{i,h},d_{i,h}])$. Note that the derivative of sigmoid function $\nu_{R,i}(x)$ equals $A_i\lambda_i e^{-\lambda_i(x-m_i)}/(1+e^{-\lambda_i(x-m_i)})^2$ with a global maximum value of $A_i\lambda_i/4$ attained at $x=m_i$. For example, Fig.3(b) displays the derivative $6e^{-3(x-4)}/(1+e^{-3(x-4)})^2$ of sigmoid function $\nu_R(x) = 2/(1+e^{-3(x-4)})$ shown in Fig.3(a). We reformulate the expression for σ_1 (in section 2.3), next.

$$\sigma_1(\boldsymbol{F},\boldsymbol{E}) = \int_0^1 \frac{\sum_{i=1}^{N}[\nu_{R,i}(\theta_{R,i}(c_{i,h})) + \nu_{R,i}(d_{i,h})]}{\sum_{i=1}^{N}[\nu_{R,i}(\theta_{R,i}(a_{i,h} \wedge c_{i,h})) + \nu_{R,i}(b_{i,h} \vee d_{i,h})]} dh =$$

$$= \int_0^1 \frac{\sum_{i=1}^{N}[A_i + Z_{R,i}([c_{i,h},d_{i,h}])]}{\sum_{i=1}^{N}[A_i + Z_{R,i}([a_{i,h},b_{i,h}] \vee [c_{i,h},d_{i,h}])]} dh = \int_0^1 \frac{A + Z_{\Delta^N}(\boldsymbol{E}(h))}{A + Z_{\Delta^N}(\boldsymbol{F}(h) \vee \boldsymbol{E}(h))} dh \text{, where}$$

$A = A_1 + \ldots + A_N$, moreover $Z_{\Delta^N}(\boldsymbol{E}(h)) = Z_{R,1}([c_{1,h},d_{1,h}]) + \ldots + Z_{R,N}([c_{N,h},d_{N,h}])$ in the numerator, and likewise in the denominator. Occam razor semantics is interpreted next.

Let $\boldsymbol{E}_l \in \mathsf{F}^N$, $l \in \{1,\ldots,L\}$ be granules (i.e. rule $\boldsymbol{E}_l \rightarrow c_l$ antecedents) competing over an input granule $\boldsymbol{F}_0$, i.e. the largest $\sigma(\boldsymbol{F}_0 \leq \boldsymbol{E}_l)$ is sought. It follows that winner $\boldsymbol{E}_J$ among granules $\boldsymbol{E}_1,\ldots,\boldsymbol{E}_L$ will be the one whose size needs to be modified, comparatively, *the least* (over all $h \in (0,1]$) so as to "barely" include $\boldsymbol{F}_0$. In the aforementioned sense, winner granule $\boldsymbol{E}_J$ is the simplest hypothesis that fits the data; that is the meaning of *Occam razor* semantics here. A similar interpretation also holds for inclusion measure σ_2 (in section 2.3).

The (granular) FLR supports two different modes of reasoning, namely *Generalized Modus Ponens* and *Reasoning by Analogy* [21]. More specifically, on one hand, *Generalized Modus Ponens* is a common form of deductive reasoning whereby, in the context of this work, given both a rule $\boldsymbol{E}_l \rightarrow c_l$, $l \in \{1,\ldots,L\}$, and an antecedent granule $\boldsymbol{F}_0$ such that $\boldsymbol{F}_0 \leq \boldsymbol{E}_l$ it follows c_l; hence, generalized modus ponens is directly supported by the *granular FLR*. On the other hand, *Reasoning by Analogy* is a mode of approximate reasoning suitable for dealing with incomplete knowledge. More specifically, given a set of rules $\boldsymbol{E}_l \rightarrow c_l$, $l=1,\ldots,L$ as well as an antecedent granule $\boldsymbol{F}_0$, such that $\boldsymbol{F}_0 \leq \boldsymbol{E}_l$ holds for no $l \in \{1,\ldots L\}$, the *granular FLR* classifier selects the rule which best fits the data ($\boldsymbol{F}_0$) in the Occam razor sense explained above. An alternative interpretation is presented next.

Inclusion measure $\sigma(\boldsymbol{F}_0 \leq \boldsymbol{E}_l)$ can be interpreted as the degree of truth of implication "$\boldsymbol{F}_0 \Rightarrow \boldsymbol{E}_l$" involving the truth values $\boldsymbol{F}_0$ and $\boldsymbol{E}_l$, respectively, of two propositions. Note that various mechanisms have been proposed for calculating a degree of truth of implication "$\boldsymbol{F}_0 \Rightarrow \boldsymbol{E}_l$" given the truth values $\boldsymbol{F}_0$ and $\boldsymbol{E}_l$ [30]. The novelty here is that the truth values $\boldsymbol{F}_0$ and $\boldsymbol{E}_l$ of the two propositions involved in implication "$\boldsymbol{F}_0 \Rightarrow \boldsymbol{E}_l$" take values in a general complete lattice [14] rather than taking values solely in the unit interval [0,1]. However, the truth of implication "$\boldsymbol{F}_0 \Rightarrow \boldsymbol{E}_l$" here takes values in the unit interval [0,1]. In particular, we define the truth of implication "$\boldsymbol{F}_0 \Rightarrow \boldsymbol{E}_l$" to be equal to $\sigma(\boldsymbol{F}_0 \leq \boldsymbol{E}_l)$. Therefore, the *granular FLR* classifier carries out tunable inferences. As a special case consider the truth table of the implication function "$A \Rightarrow B$" in Table 1, where both A and B take on (binary) values in the set {0,1}. Apparently, the inclusion measure function "$\sigma(A \leq B)$" is identical to function "$A \Rightarrow B$". In conclusion, the inclusion measure function "$\sigma(A \leq B)$" can be interpreted as the degree of truth of implication "$A \Rightarrow B$". Next, we compute the complexity of *granular FLR*.

When a data pair $(\boldsymbol{F}_i, c_i) \in \mathsf{F}^N \times \mathsf{C}$, i=1,...,$n$ is presented for training then the fuzzy inclusion measure $\sigma(\boldsymbol{F}_i \leq \boldsymbol{E}_l)$, l=1,…,L is calculated for all L granules $\boldsymbol{E}_l$ in *RB*. The worst-case training scenario is to keep "resetting" all L pairs in *RB* for every input. Since both 1) the largest value for L is L=n, and 2) a single pass through the data suffices for learning, it follows that the *training complexity* is quadratic $\mathsf{O}(n^2)$ in the number n of the data for training. Likewise, it can be shown that the *testing complexity* of the FLR classifier is linear $\mathsf{O}(n)$.

3.3. Comparative discussion

The *granular FLR* carries out *lattice computing*, or *LC* for short, where LC was (roughly) defined as lattice-theory-based Computational Intelligence [15]. More accurately, LC is defined as an evolving collection of tools and methodologies that can process disparate types of data including logic values, numbers, sets, symbols, and graphs based on mathematical lattice theory with emphasis on clustering, classification, regression, pattern analysis, and knowledge representation applications. LC is currently used by different authors in various domains including Logic and Reasoning [53], Mathematical Morphology [37], Computational Intelligence [19], and Formal Concept Analysis [13].

There are inherent similarities as well as substantial differences between FAM and the *granular FLR*. For instance, both FAM and *granular FLR* learn rapidly in a single pass through the training data by applying, in principle, the same algorithm. A cluster computed by either algorithm corresponds to the core of a fuzzy set. Nevertheless, a cluster for *granular FLR* is a N-tuple FIN in F^N, including N-dimensional hyperboxes, the latter are the only type of clusters computable by FAM solely in the unit hypercube. Moreover, only the (*granular*) *FLR* can deal "in principle" with "missing" data and/or "don't care" data in a constituent

(complete) lattice by replacing the aforementioned data, respectively, by the *least* and the *greatest* element *O* and *I* in the corresponding constituent lattice [26].

The *granular FLR* is interpreted as a reasoning scheme, which supports two different modes of reasoning, namely *Generalized Modus Ponens* and *Reasoning by Analogy* as it was explained above; whereas, FAM makes a decision based on an "objective function" such as a *Choice* (*Weber*) function or a *Match* function.

A substantial advantage of the *granular FLR* is its capacity to optimally tune either a strictly increasing (positive valuation) function $v(x)$ or a (bijective) strictly decreasing function $\theta(x)$ in a data dimension in the Euclidean space R^N; whereas FAM uses, implicitly as well as quite restrictively, only $v(x)=x$ and $\theta(x)=1-x$ in a data dimension in the unit hypercube. A couple of *granular FLR* drawbacks are presented next.

FAM's *proliferation problem*, regarding clusters, is inherited to *granular FLR*. However, the *granular FLR* is equipped with such tools as a tunable inclusion measure as well as a tunable metric function to reduce "in principle" the number of hyperboxes. Another drawback of the *granular FLR*, also inherited from FAM, is that the learned clusters (in particular their total number, size, and location) depend on the order of presenting the training data. A potential solution is to employ an ensemble of *granular FLR*s in order to boost performance [19].

4. An industrial prediction application

This section demonstrates prediction-by-classification in an industrial application.

4.1. The physical problem

The Phosphoric Fertilizers Industry (PFI) in N. Karvali, Greece produces industrial fertilizer by spraying Ammonium Nitrate (AN) solution on small solid particles inside a rotating *pan granulator* mill [23]. The end-product consists of small fertilizer *granules* each having size in the range of a few millimeters (Fig. 4). High quality fertilizer specifications demand at least 95% of the granules to be in the range 2-5 mm having as spherical shape as possible; moreover, fertilizer granules should be covered "uniformly" with Ammonium Nitrate (AN).

In order to retain a "granular" fertilizer product during both storage and transportation, the *stiffness* of fertilizer granules should be above a threshold value. Hence, the aforementioned stiffness is sampled during production and, if necessary, corrective control actions are taken. However, a problem arises as the measurement of stiffness takes time (around one hour) since it is carried out mechanically in the lab using a small "vibrating" cylinder including eight sieves (with successively increasing grid size from 1 mm up to 4.5 mm). Therefore, in the

mean time, a good deal of fertilizer product may go to waste. It is of interest to develop a reliable predictive model for stiffness towards taking corrective control actions much sooner.

The development of a "first principles" prediction model was phased out due to the inherent complexity of the industrial process. Instead, we focused our efforts on inducing a predictive classification model from real-world measurements – Note that prediction by classification is a common practice in machine learning [19]. More specifically, our objective here is to induce a stiffness class label in the set {"small", "medium", "large"} from populations of measurements obtained by digital image processing techniques as explained next.

4.2. Data acquisition

An "unbiased" *population* sample including a variable number of up to a few hundred fertilizer granules was collected every 8 hours in March 2005 for 14 consecutive days during *steady state* production of fertilizer type CaN27. An aforementioned population was spread on a black background at a specific distance from an image grabber; in conclusion, grayscale digital images were obtained under white light. The *area* (in mm^2) as well as an index of *circularity* (in the interval [0,1]) were computed for each fertilizer granule by digital image processing techniques [11] as follows.

The images were thresholded using histogram processing techniques; hence, binary images were produced. Both the *area* and *perimeter* of each fertilizer granule in an image were computed. For each fertilizer granule, a diameter d_a was computed as the diameter of an equal area circle. Another diameter d_p was computed for each granule as the diameter of an equal perimeter circle. Ratio d_a/d_p was a granule's *circularity* index CI ≤ 1.

In addition, the "Ammonium Nitrate (AN)" sprayed on a fertilizer granule was quantified by the grayscale *brightness difference* under white light (Fig. 4(a)) from under ultraviolet light (Fig. 4(b)). Hence, for each fertilizer population (sample), one distribution of measurements was produced per variable *area*, *circularity*, and *brightness difference*. The latter three distributions represented the corresponding digital image(s).

Compared to the mechanical measurement technique described in section 4.1, the proposed digital image processing techniques were clearly superior for several reasons. First, as soon as a population (sample) of fertilizer granules becomes available, digital image processing techniques were at least 10 times faster than the aforementioned mechanical measurement technique. Second, the resolution of measurements by the former techniques was at least 2 orders of magnitude finer than the 0.5 mm resolution of mechanical measurement. Third, mechanical measurement is questionable because longer fertilizer granules tend to slip through a sieve, hence actual fertilizer granule size tends to be (slightly) larger than what it is

measured mechanically. Fourth, another advantage for the digital image processing techniques is that the circularity index CI of individual fertilizer granules can be measured.

Fertilizer granule *stiffness* was measured manually (hence slowly) for each granule in a population as follows. We measured the peak force (in the range 10 N to 40 N) required to crush a fertilizer granule using a special force-meter device. However, in the context of this work, we did not use a distribution of *stiffness* measurements. Instead, an expert from the industry classified a *stiffness* distribution in one of the three class labels "small", "medium", and "large". In conclusion, our data consisted of 42 triplets of distributions, each triplet was given together with its corresponding class (*stiffness*) label.

4.3. FIN representation and interpretation issues

A popular representation of a distribution (of real number measurements) is by a histogram. For practical reasons one can claim that when the number of samples in a distribution becomes "large enough" then a histogram goes to a probability density function (pdf). Due to a bijection (i.e. one-to-one correspondence) between pdfs and CDFs (cumulative distribution functions) there follows a bijection between histograms and CDFs. In the following we show representation of a CDF by a FIN.

It is well known that a FIN F: $(0,1]\rightarrow\tau_O(\mathsf{R})$ can represent a CDF $P(x)$ as follows: $m_F(x)=2P(x)$ for $x\leq x_0$, furthermore $m_F(x)=2[1-P(x)]$ for $x\geq x_0$, where $P(x_0)=0.5$ and m_F: $\mathsf{R}\rightarrow(0,1]$ is the membership function of FIN F [18], [19], [24]. Hence, a FIN represents a histogram.

In practice, a FIN can be computed from a distribution of measurements by algorithm CALFIN [18], [19], [23] for any number of samples in a distribution. A statistical interpretation of a FIN follows. If a "large" number of samples is drawn independently according to a pdf $p_0(x)$ and a *FIN* F is constructed by algorithm CALFIN then interval $F(h)$ constitutes an *interval of confidence at level-h* in the sense that a random number drawn according to $p_0(x)$ is expected to fall 1) *inside* interval $F(h)$ with probability $100(1-h)\%$, and 2) either *below* or *above* interval $F(h)$ with probability $50h\%$ [18], [19], [23], [24]. For graphical illustrations regarding the computation of a FIN by algorithm CALFIN the interested reader may refer to [18], [24], [35]. A subtle practical advantage of using a FIN (computed by algorithm CALFIN) is explained next.

A popular practice in the literature for representing a distribution of measurements is using up to second order statistics including both the corresponding average and standard deviation. However, the aforementioned practice is error-prone because potentially important "higher order statistics", such as skewness, etc. are ignored. Note that the need to employ higher-order statictics was already acknowledged in practice including climate modeling [5], [36].

Different authors have already proposed new techniques, including the *ordered weighted averaging* (OWA), for considering certain statistics [54]. Due to the aforementioned bijection between FINs and CDFs, a comparative advantage of a FIN is that a FIN can consider statistics of all orders [19], [23], [24].

A FIN can represent a fuzzy number with an arbitrary membership function. It is noteworthy that other authors have also pursed an induction of fuzzy sets with arbitrary membership functions using different lattice theory techniques based on fuzzy logic [52]. A comparative advantage here is the capacity to introduce tunable nonlinearities towards improving performance as demonstrated below using stochastic search optimization techniques.

4.4. Computational experiments and results

In our computational experiments, described in this section, we employed 42 labeled (3-tuple) FIN data induced from industrial digital images as detailed above. The data were numbered in the order recorded. We used the first 25 data for training, the next 10 data for validation, and the last 7 data for testing. Table 2 displays the distribution of the aforementioned data for training/validation/testing in three categories labeled **small**, **medium**, and **large** (*stiffness*).

In our experiments we confirmed that the *granular FLR* always learned fast in a single pass through the training data, as expected. We employed one (sigmoid) positive valuation function $v_{R,i}(x) = A_i /(1+e^{-\lambda_i (x-m_i)})$ with three parameters A_i, λ_i, m_i per (data) dimension in three dimensions i=1,2,3. An additional parameter included the granule threshold size Z_{crit} (Fig. 1). In conclusion, we pursued optimal estimation of 10 parameters by a genetic algorithm (GA) immediately after training the *granular FLR*, as explained next.

We represented an individual GA solution of the *granular FLR* using 10 parameters. Hence, the chromosome of an individual solution consisted of 10 genes. Each gene used 16 bits to encode a single parameter value. In conclusion, a chromosome was 160 bits long. The population of the genetic algorithm included 100 individuals. The genetic algorithm employed multipoint crossover and roulette wheel selection for reproduction, elitism, multipoint mutation, and adaptive crossover-mutation rates. Genetic optimization was enhanced by specialized operators including both the Adaptive SEarch space Range (ASER) and the microgenetic one [24]. Objective function $f(P_{trn},P_{val},Q;w,\varepsilon)= wP_{trn}+(1-w)P_{val}+\varepsilon/Q$ calculated the fitness of an individual (solution), where P_{trn} is the "training data" classification accuracy, P_{val} is the "validation data" classification accuracy, Q is the total number of granules (clusters), w is a parameter in the open interval (0,1), and ε is a small positive parameter. Note that we chose the aforementioned form of objective function $f(P_{trn},P_{val},Q;w,\varepsilon)$ empirically. In particular, we chose w=0.1 so as to consider both the training- and the validation- data with

emphasis on the validation data; furthermore, we chose a small value ε=0.001 so as to break "ties" in favor of solutions with smaller Q. Evolution terminated when the fitness of the elite individual (solution) did not improve for 100 generations in a row.

Table 3 displays the training data parameter ranges, which we fed the GA with as well as the corresponding optimally estimated parameter values. We point out that the data in our experiments were not normalized. The optimally estimated sigmoid functions are displayed, over a domain of interest, in Fig. 5 (a), (b), and (c). It is remarkable that a positive valuation in Fig.5 changes by a different amount over a domain of interest. The latter signifies the "discriminatory capacity" of the corresponding input variable as explained below.

Using the optimal parameter estimates of Table 3, the *granular FLR* induced one granule (cluster) per class. Each one of the latter clusters can be interpreted as a rule for classification (Fig. 5). For instance, rule R1 can be interpreted as

If "*area* is A1".AND."*circularity* is C1".AND."*brightness difference* is BD1" then "fertilizer granule *stiffness* is **small**"

Note that a FIN, plotted in bold in Fig. 5, is the "envelope" of all the FINs it contains.

The testing data were applied once, after training. We recorded a testing data classification accuracy of 100% as shown in the first line of Table 4.

We remark that "envelope" FIN granules A1, A2, and A3 in Fig. 5 in rules R1, R2, and R3, respectively, show that all the corresponding training data FINs clearly lie inside the saturated region of the corresponding sigmoid positive valuation function shown in Fig. 5(a). In other words, the latter positive valuation function essentially remains constant over the training data domain. Hence, we considered the following hypothesis: "input variable *area* is redundant and it can be omitted". We confirmed the aforementioned hypothesis by additional computational testing experiments using only input variables *circularity* and *brightness difference*. A 100% testing data classification accuracy resulted in, again.

We repeated the testing phase using positive valuation function $v(x)=x$ in every data dimension. In the latter case we recorded one misclassification as shown in the second line of Table 4. For a further comparison, we carried out a number of additional *k-nearest neighbor* (kNN) classification experiments for k=1 as follows.

First, we applied kNN in metric space F^3, 1) using the three sigmoid positive valuation functions whose (optimal) parameters are shown in Table 3, and 2) using positive valuation function $v(x) = x$ in a data dimension. Second, we replaced a distribution (of measurements) by its corresponding median. Then, we applied kNN in the Euclidean space R^3, 1) using the three sigmoid positive valuation functions whose (optimal) parameters are shown in Table 3,

and 2) using positive valuation function $v(x) = x$ in a data dimension. In each aforementioned experiment we recorded one misclassification as shown in Table 4 lines 3-6, respectively.

4.5. Discussion of the results

The *granular FLR* here processed N-tuples of histograms, which represented digital images. A histogram was represented by a non-parametric FIN. FINs are preferable in practice because they can represent data statistics of all orders [18], [22], [23], [24], [35].

This work has demonstrated the advantage of introducing tunable nonlinearities resulting in a testing data classification accuracy of 100% as summarized comparatively in Table 4. However, *granular FLR* required significantly more time for calculating "genetically" optimal sigmoid function parameter values.

Another advantage of the *granular FLR* is that it induced descriptive decision-making knowledge (rules) from the training data, whereas a kNN classifier can not produce any rules.

Previous work has employed "hyperbolic tangent" positive valuation functions [24], nevertheless sigmoid ones are preferable in the context of this work for the following reasons. A sigmoid function $v_R(x)= A/(1+e^{-\lambda(x-m)})$ varies between 0 and $A/2>0$, therefore it is suitable for defining an inclusion measure function (σ) since the latter (function) needs to be zero for the *least* interval (see Definition A.9 in the Appendix). Another advantage is that a sigmoid positive valuation function, with optimally estimated parameters A, λ, and m, can avoid conventional "data preprocessing" (towards data normalization). Moreover, a sigmoid function is popular in neural computing. This work has also presented experimental evidence that a (sigmoid) positive valuation can suggest input variable selection as explained next.

When a positive valuation function practically remains constant over the domain of an input variable, e.g. input variable *area* in Fig. 5, then and only then we conjecture that the aforementioned variable is not significant and could be omitted. The aforementioned conjecture, namely here *null hypothesis H0*, remains to be thoroughly tested statistically in a future work since the problem of input variable selection may be significant in practice [32].

Additional advantages include *granular FLR*'s capacity to deal "in principle" with both "missing" and "don't care" data as well as its capacity to support two modes of reasoning, namely *Generalized Modus Ponens* and *Reasoning by Analogy*.

5. Conclusion

This work has introduced the *granular FLR* as an extension of FAM neural classifier for processing granular FIN data including (fuzzy) numbers, intervals, and cumulative distribution functions. Based on a fundamentally improved novel mathematical notation, FINs can directly be associated with α-cuts of fuzzy numbers towards a wider proliferation of our proposed techniques in practice including fuzzy system applications. Section 4.5 summarized comparatively a number of advantages regarding the *granular FLR*.

Experimental results have confirmed the viability of our proposed techniques. In particular, the effectiveness of the *granular FLR* classifier was demonstrated here comparatively in an industrial prediction-by-classification application.

Future plans include extensive statistical testing regarding, first, the general effectiveness of our proposed techniques and, second, the *null hypothesis H0* (see in section 4.5) towards an effective input variable selection. Future work also includes a further (linguistic) extension of the *granular FLR* towards a sound accommodation of interval-FINs.

Appendix

This Appendix summarizes basic lattice theory notions and useful results in order to make this work "self-contained". A novel theorem, as well as its proof, is also shown.

A.1. Posets

A fundamental abstract notion is "(partial) order" [4] defined next.

Definition A.1. Let P be a nonempty set. A (*partial*) *order* ($\leq$) is a binary relation $\leq \subseteq P \times P$, which satisfies 1) $x \leq x$ (Reflexivity); 2) $x \leq y$ and $y \leq x \Rightarrow x = y$ (Antisymmetry); 3) $x \leq y$ and $y \leq z \Rightarrow x \leq z$ (Transitivity).

A partial order implies a *partially ordered set*, or *poset* for short, as follows.

Definition A.2. A *poset* is a pair $(P,\leq)$, where P is a set and $\leq$ is a *partial order* relation in P.

We remark that if $x \leq y$ (or, $y \leq x$) then x and y are called *comparable*; otherwise, x and y are called *incomparable* or *parallel*, symbolically $x \| y$, for $x,y \in P$. A poset $(P,\leq)$, which includes solely comparable elements is called *totally-ordered* or, equivalently, *chain*. If $x \leq y$ and $x \neq y$ in a poset $(P,\leq)$ then one writes $x < y$.

A poset can contain at most one element O, which satisfies $O \leq x$ for all $x \in P$. Such an element, if it exists, is called *least* element of P. The dual *greatest* element of P is denoted by I.

By "a covers b" in a poset $(P,\leq)$ one means that $b < a$ but $b < x < a$ for no $x \in P$. Let $(P,\leq)$ be a poset with least element O. Every $x \in P$ which covers O, if such x exists, is called *atom*.

An *interval* $[a,b]$, in a poset $(P,\leq)$, is defined as the set $[a,b] \doteq \{x \in P: a \leq x \leq b\}$. Let $\tau(P)$ denote the set of intervals in $(P,\leq)$. It follows that a trivial interval $[x,x] \in \tau(P)$ is an atom in poset $(\tau_O(P) = \tau(P) \cup \{O\},\leq)$, the latter is ordered by set inclusion with least element (empty set) O.

Definition A.3. Let $(P,\leq)$ and $(Q,\leq)$ be posets. A map $\psi: P \to Q$ is called 1) *Order-preserving* (or, simply, *monotone*) if $x \leq y$ in $(P,\leq)$ implies $\psi(x) \leq \psi(y)$ in $(Q,\leq)$, and 2) *Order-isomorphism* (or, simply, *isomorphism*) if both "$x \leq y$ in $(P,\leq) \Leftrightarrow \psi(x) \leq \psi(y)$ in $(Q,\leq)$" and "ψ is onto Q".

We remark that when there is an isomorphism between two posets $(P,\leq)$ and $(Q,\leq)$ then the aforementioned posets are called *isomorphic*, symbolically $(P,\leq) \cong (Q,\leq)$.

The *dual* of a poset $(P,\leq)$ is another poset defined by the converse order relation, symbolically $(P,\leq)^{\partial} \equiv (P,\leq^{\partial}) \equiv (P,\geq)$.

A.2. Crisp lattices

Lattice theory was compiled creatively by Garrett Birkhoff [4]. A (mathematical) lattice can be defined based on a poset as follows.

Let $(P,\leq)$ be a poset and $X\subseteq P$. An *upper bound* of X is a $a\in P$ with $x\leq a$, $\forall x\in X$. The *least upper bound* (l.u.b.), if it exists, is the unique upper bound contained in every upper bound. The l.u.b. is also called *supremum* or *lattice join* of X and denoted by $\mathtt{sup}X$ or $\vee X$. The notions *lower bound* of X and *greatest lower bound* (g.l.b.) of X are defined dually. The g.l.b. is also called *infimum* or *lattice meet* of X and denoted by $\mathtt{inf}X$ or $\wedge X$.

Definition A.4. A *lattice*, or equivalently *crisp lattice*, is a poset $(\mathsf{L},\leq)$ any two of whose elements have both a g.l.b., denoted by $x\wedge y$, and a l.u.b., denoted by $x\vee y$, in L.

A lattice $(\mathsf{L},\leq)$ is called *complete* when a subset $X\subseteq\mathsf{L}$ has both a l.u.b. and a g.l.b. in L. Setting $X=\mathsf{L}$ it follows that a nonvoid complete lattice contains both a *least* element and a *greatest* element denoted, respectively, by O and I. In this work we deal solely with complete lattices.

A *sublattice* of a lattice $(\mathsf{L},\leq)$ is another lattice $(\mathsf{S},\leq)$ such that $\mathsf{S}\subseteq\mathsf{L}$. The *dual* of a lattice $(\mathsf{L},\leq)$, symbolically $(\mathsf{L},\leq)^{\partial}\equiv(\mathsf{L},\leq^{\partial})\equiv(\mathsf{L},\geq)$, is a lattice. An *atomic lattice* $(\mathsf{L},\leq)$ is one in which every element is a joint of atoms.

Definition A.5. A *positive valuation* in a lattice $(\mathsf{L},\leq)$ is a real function v: $\mathsf{L}\rightarrow\mathsf{R}$, which satisfies both $v(x)+v(y)=v(x\wedge y)+v(x\vee y)$ and $x<y\Rightarrow v(x)<v(y)$.

A positive valuation v: $\mathsf{L}\rightarrow\mathsf{R}$ in a lattice $(\mathsf{L},\leq)$ implies a *metric* function d: $\mathsf{L}\times\mathsf{L}\rightarrow\mathsf{R}_0^+$ given by

$$d(a,b)=v(a\vee b)-v(a\wedge b) \tag{E1}$$

For convenience, the definition of "metric" is shown next.

Definition A.6. A *metric* in a set A is a nonnegative real function d: $A\times A\rightarrow\mathsf{R}_0^+$, which satisfies 0) $d(x,y)=0\Rightarrow x=y$; 1) $d(x,x)=0$; 2) $d(x,y)=d(y,x)$; 3) $d(x,z)\leq d(x,y)+d(y,z)$ (Triangle Inequality).

If only conditions 1) to 3) are satisfied in Definition A.6 then d is called *pseudo-metric*. A *metric space* is a pair (A,d) including both a set A and a metric d: $A\times A\rightarrow\mathsf{R}_0^+$.

Of particular interest here is a Cartesian product lattice $(\mathsf{L},\leq)=(\mathsf{L}_1,\leq_1)\times\ldots\times(\mathsf{L}_N,\leq_N)$ of *constituent* lattices $(\mathsf{L}_i,\leq_i)$, i=1,…,N. Note that each constituent lattice $(\mathsf{L}_i,\leq_i)$ is characterized by a specific order relation $\leq_i$, i=1,…,N. Nevertheless, a general lattice order will be denoted here by the same symbol $\leq$, in the interest of simplicity. On one hand, *positive valuation*

functions v_i: $\mathsf{L}_i \to \mathsf{R}$ in constituent lattices $(\mathsf{L}_i,\leq)$, i=1,…,N imply both a positive valuation v: $\mathsf{L} \to \mathsf{R}$ given by $v(x_1,\ldots,x_N)= v_1(x_1)+\ldots+v_N(x_N)$ and a *Minkowski metric* d_p: $\mathsf{L} \to \mathsf{R}_0^+$ given by

$$d_p(\mathbf{x},\mathbf{y}) = \left[d_1^p(x_1,y_1) + \ldots + d_N^p(x_N,y_N) \right]^{1/p}, \text{ where } \mathbf{x}= (x_1,\ldots,x_N),\ \mathbf{y}=(y_1,\ldots,y_N),\ p \in \mathsf{R} \qquad \text{(E2)}$$

On the other hand, *isomorphic* functions θ_i: $\mathsf{L}_i \to \mathsf{L}_i$ from a constituent lattice $(\mathsf{L}_i,\leq)$ to its dual $(\mathsf{L}_i,\geq)$, i=1,…,N imply an isomorphic function θ: $\mathsf{L} \to \mathsf{L}$ from lattice $(\mathsf{L}= \mathsf{L}_1 \times \ldots \times \mathsf{L}_N,\leq)$ to its dual $(\mathsf{L},\geq)$ given by $\theta(x_1,\ldots,x_N)= (\theta_1(x_1),\ldots,\theta_N(x_N))$.

Poset $(\tau_O(\mathsf{L})= \tau(\mathsf{L}) \cup \{O\},\leq)$ of intervals, ordered by set inclusion, is an *atomic lattice* with least element O.

Based on a positive valuation v: $\mathsf{L} \to \mathsf{R}$, the *size* of an interval in $(\tau(\mathsf{L}),\leq)$ is defined as follows.

Definition A.7. The *size* Z_L: $\tau(\mathsf{L}) \to \mathsf{R}_0^+$ of an interval $[a,b] \in \tau(\mathsf{L})$ in lattice $(\mathsf{L},\leq)$, with respect to a positive valuation v: $\mathsf{L} \to \mathsf{R}$, is defined as $Z_\mathsf{L}([a,b]) = d(a,b) = v(b) - v(a)$.

A.3. Fuzzy lattices

A *fuzzy set* is denoted here by a pair (U,m), where U is a *universe of discourse* and m is a *membership* function m: $U \to [0,1]$. A *fuzzy lattice* [19] fuzzifies the binary relation "≤"in a crisp lattice as follows.

Definition A.8. A *fuzzy lattice* is a triple $(\mathsf{L},\leq,m)$, where $(\mathsf{L},\leq)$ is a crisp lattice and $(\mathsf{L} \times \mathsf{L},m)$ is a fuzzy set such that $m(x,y)=1$ if and only if $x \leq y$.

Function m: $U \to [0,1]$ in Definition A.8 is a *weak* (*fuzzy*) *partial order* relation in the sense that both $m(x,y)=1$ and $m(y,z)=1$ imply $m(x,z)=1$, whereas if either $m(x,y) \neq 1$ or $m(y,z) \neq 1$ then $m(x,z)$ could be any number in the interval [0,1]. A (complete) lattice can be fuzzified by an *inclusion measure* function defined next.

Definition A.9. Let $(\mathsf{L},\leq)$ be a complete lattice with least element O. An *inclusion measure* is a function σ: $\mathsf{L} \times \mathsf{L} \to [0,1]$, which satisfies the following conditions 0) $\sigma(x,O)=0$, $x \neq O$; 1) $\sigma(x,x) = 1$, $\forall x \in \mathsf{L}$; 2) $x \wedge y < x \Rightarrow \sigma(x,y) < 1$; and, 3) $u \leq w \Rightarrow \sigma(x,u) \leq \sigma(x,w)$ (Consistency Property).

We remark that $\sigma(x,y)$ may be interpreted as a (fuzzy) degree of inclusion of x in y. Therefore, notations $\sigma(x,y)$ and $\sigma(x \leq y)$ are used interchangably. It turns out that if σ: $\mathsf{L} \times \mathsf{L} \to [0,1]$ is an inclusion measure then $(\mathsf{L},\leq,\sigma)$ is a fuzzy lattice [19], [21]. A couple inclusion measures can be defined in a complete lattice based on a positive valuation function as shown next [19].

Theorem A.10. If v: $\mathsf{L}\rightarrow\mathsf{R}_0^+$ is a positive valuation in a complete lattice $(\mathsf{L},\leq)$, with $v(O)=0$, then both functions $k(x,u) = v(u)/v(x\vee u)$ and $s(x,u) = v(x\wedge u)/v(x)$ are inclusion measures.

We point out that Theorem A.10 calls for a *nonnegative* positive valuation function v.

Proposition A.11. Let function σ_i: $\mathsf{L}_i\times\mathsf{L}_i\rightarrow[0,1]$ be an inclusion measure in a constituent lattice $(\mathsf{L}_i,\leq)$, i=1,…,N. Then, *convex combination* $\sigma(\boldsymbol{x},\boldsymbol{y}) = \lambda_1\sigma_1(x_1,y_1)+\ldots+\lambda_N\sigma_N(x_N,y_N)$ is an inclusion measure in the Cartesian product lattice $(\mathsf{L},\leq)= (\mathsf{L}_1,\leq)\times\ldots\times(\mathsf{L}_N,\leq)$.

Proof

By "convex combination" above we mean a set $\lambda_1,\ldots,\lambda_N$ of positive numbers such that $\lambda_1+\ldots+\lambda_N = 1$. Next, we show that function $\sigma(\boldsymbol{x},\boldsymbol{y}) = \lambda_1\sigma_1(x_1,y_1)+\ldots+\lambda_N\sigma_N(x_N,y_N)$ satisfies the conditions of Definition A.9.

0) $\sigma(\boldsymbol{x},\boldsymbol{O}) = \lambda_1\sigma_1(x_1,O)+\ldots+\lambda_N\sigma_N(x_N,O) = 0$, $\boldsymbol{x}\neq\boldsymbol{O}=(O,\ldots O)$;

1) $\sigma(\boldsymbol{x},\boldsymbol{x}) = \lambda_1\sigma_1(x_1,x_1)+\ldots+\lambda_N\sigma_N(x_N,x_N) = \lambda_1+\ldots+\lambda_N = 1$, $\forall x_i\in\mathsf{L}_i$;

2) $\boldsymbol{x}\wedge\boldsymbol{y} < \boldsymbol{x} \Rightarrow \exists i\in\{1,...,N\}: x_i\wedge y_i < x_i \Rightarrow \sigma(x_i,y_i) < 1$; Hence,

   $\boldsymbol{x}\wedge\boldsymbol{y} < \boldsymbol{x} \Rightarrow \sigma(\boldsymbol{x},\boldsymbol{y}) = \lambda_1\sigma_1(x_1,y_1)+\ldots+ \lambda_N\sigma_N(x_N,y_N) < \lambda_1+\ldots+\lambda_N = 1$;

3) $\boldsymbol{u}\leq\boldsymbol{w} \Rightarrow u_i\leq w_i$, $\forall i\in\{1,...,N\} \Rightarrow \sigma(x_i,u_i)\leq\sigma(x_i,w_i)$; Hence,

   $\boldsymbol{u}\leq\boldsymbol{w} \Rightarrow \lambda_1\sigma_1(x_1,u_1)+\ldots+\lambda_N\sigma_N(x_N,u_N)\leq\lambda_1\sigma_1(x_1,w_1)+\ldots+\lambda_N\sigma_N(x_N,w_N) \Rightarrow \sigma(\boldsymbol{x},\boldsymbol{u})\leq\sigma(\boldsymbol{x},\boldsymbol{w})$.

The latter completes the proof of Proposition A.11.

References

[1] A. Al-Daraiseh, A. Kaylani, M. Georgiopoulos, M. Mollaghasemi, A.S. Wu, and G. Anagnostopoulos, GFAM: evolving fuzzy ARTMAP neural networks, Neural Networks 20 (8) (2007) 874-892.

[2] A. Bargiela and W. Pedrycz, Granular Computing: An Introduction (Kluwer Academic Publishers, Boston, The Kluwer International Series in Engineering and Computer Science 717, 2003).

[3] A. Bargiela, W. Pedrycz, and K. Hirota, Granular prototyping in fuzzy clustering, IEEE Transactions on Fuzzy Systems 12 (5) (2004) 697-709.

[4] G. Birkhoff, Lattice Theory (AMS, Providence, Colloquium Publications 25, 1967).

[5] B.K. Biswas, Y.M. Svirezhev, and B.K. Bala, A model to predict climate-change impact on fish catch in the world oceans, IEEE Transactions on Systems, Man and Cybernetics – Part A 35 (6) (2005) 773-783.

[6] G.A. Carpenter and S. Grossberg, A massively parallel architecture for self-organizing neural pattern recognition machine, Computer Vision, Graphics, and Image Undrestanding 37 (1) (1987) 54-115.

[7] G.A. Carpenter, S. Grossberg, and D.B. Rosen, Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system, Neural Networks 4 (6) (1991) 759-771.

[8] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps, IEEE Trans Neural Networks 3 (5) (1992) 698-713.

[9] J. Castro, M. Georgiopoulos, R. DeMara, and A. Gonzalez, Data-partitioning using the Hilbert space filling curves: effect on the speed of convergence of fuzzy ARTMAP for large database problems, Neural Networks 18 (7) (2005) 967-984.

[10] J. Castro, J. Secretan, M. Georgiopoulos, R. DeMara, G. Anagnostopoulos, and A.J. Gonzalez, Pipelining of fuzzy-ARTMAP without match-tracking: correctness, performance bound, and Beowulf evaluation, Neural Networks 20 (1) (2007) 109-128.

[11] V. Chatzis, V.G. Kaburlasos, and M. Theodorides, An image processing method for particle size and shape estimation, in: Proceedings of the 2nd International Scientific Conference on Computer Science, Chalkidiki, Greece, 2005.

[12] B.A. Davey and H.A. Priestley, Introduction to Lattices and Order (Cambridge University Press, Cambridge, 1990).

[13] B. Ganter and R. Wille, Formal Concept Analysis. Mathematical Foundations (Springer-Verlag, Berlin, 1999).

[14] J.A. Goguen, L-fuzzy sets, J. Math. Analysis and Applications 18 (1967) 145-174.

[15] M. Graña, Lattice computing: lattice theory based computational intelligence, in: Proceedings of the Kosen Workshop on Mathematics, Technology, and Education (MTE) 19-27, Ibaraki, Japan, 2008.

[16] S. Grossberg, Adaptive pattern classification and universal recoding, I: parallel development and coding of neural feature detectors, Biological Cybernetics 23 (1976) 121-134.

[17] S. Grossberg, Adaptive pattern classification and universal recoding, II: feedback, expectation, olfaction, and illusions, Biological Cybernetics 23 (1976) 187-202.

[18] V.G. Kaburlasos, FINs: lattice theoretic tools for improving prediction of sugar production from populations of measurements, IEEE Transactions on Systems, Man and Cybernetics – Part B 34 (2) (2004) 1017-1030.

[19] V.G. Kaburlasos, Towards a Unified Modeling and Knowledge-Representation Based on Lattice Theory (Springer, Heidelberg, Studies in Computational Intelligence 27, 2006).

[20] V.G. Kaburlasos, Neural/Fuzzy Computing Based on Lattice Theory, in: J.R. Rabual, J. Dorado, A. Pazos, eds., Encyclopedia of Artificial Intelligence (Information Science Reference, 2008) to be published.

[21] V.G. Kaburlasos, I.N. Athanasiadis, and P.A. Mitkas, Fuzzy lattice reasoning (FLR) classifier and its application for ambient ozone estimation, International Journal of Approximate Reasoning 45 (1) (2007) 152-188.

[22] V.G. Kaburlasos and A. Kehagias, Novel fuzzy inference system (FIS) analysis and design based on lattice theory. part I: working principles, International Journal of General Systems 35 (1) (2006) 45-67.

[23] V.G. Kaburlasos and A. Kehagias, Novel fuzzy inference system (FIS) analysis and design based on lattice theory, IEEE Transactions on Fuzzy Systems 15 (2) (2007) 243-260.

[24] V.G. Kaburlasos and S.E. Papadakis, Granular self-organizing map (grSOM) for structure identification, Neural Networks 19 (5) (2006) 623-643.

[25] V.G. Kaburlasos and S.E. Papadakis, Fuzzy lattice reasoning (FLR) implies a granular enhancement of the fuzzy-ARTMAP classifier, in: Joint Conference on Information Sciences (JCIS), Proceedings of the 8th International Conference on Natural Computing (NC) 1610-1616, Salt Lake City, Utah, 2007.

[26] V.G. Kaburlasos and V. Petridis, Fuzzy lattice neurocomputing (FLN) models, Neural Networks 13 (10) (2000) 1145-1170.

[27] K.H. Knuth, Lattice duality: the origin of probability and entropy. Neurocomputing 67 (2005) 245-274.

[28] T.Y. Lin, Introduction to special issues on data mining and granular computing (editorial), International Journal of Approximate Reasoning 40 (1-2) (2005) 1-2.

[29] F. Masulli and S. Rovetta, Soft transition from probabilistic to possibilistic fuzzy clustering, IEEE Transactions on Fuzzy Systems 14 (4) (2006) 516- 527.

[30] H.T. Nguyen and E.A. Walker, A First Course in Fuzzy Logic, 3rd ed. (Chapman & Hall /CRC, Boca Raton, 2005).

[31] S.E. Papadakis and V.G. Kaburlasos, Induction of classification rules from histograms, in: Joint Conference on Information Sciences (JCIS), Proceedings of the 8th International Conference on Natural Computing (NC) 1646-1652, Salt Lake City, Utah, 2007.

[32] S.E. Papadakis, P. Tzionas, V.G. Kaburlasos, and J.B Theocharis, A genetic based approach to the type I structure identification problem, Informatica 16 (3) (2005) 365-382.

[33] W. Pedrycz, Knowledge-Based Clustering – From Data to Information Granules (John Wiley & Sons, Hoboken, 2005).

[34] L.F.C. Pessoa and P. Maragos, Neural networks with hybrid morphological/rank/linear nodes: a unifying framework with applications to handwritten character recognition, Pattern Recognition 33 (6) (2000) 945-960.

[35] V. Petridis, and V.G. Kaburlasos, FINkNN: a fuzzy interval number k-nearest neighbor classifier for prediction of sugar production from populations of samples, Journal of Machine Learning Research, 4(Apr) (2003) 17-37.

[36] D.W. Pierce, Beyond the means: validating climate models with higher-order statistics, Computing in Science & Engineering 6(5) (2004) 22-29.

[37] G.X. Ritter and J.N. Wilson, Handbook of Computer Vision Algorithms in Image Algebra, 2nd ed. (Boca Raton, Florida, CRC Press, 2000).

[38] G.X. Ritter, J.L. Diaz-de-Leon, and P. Sussner, Morphological bidirectional associative memories, Neural Networks 12 (6) (1999) 851-867.

[39] G.X. Ritter, P. Sussner, and J.L. Diaz-de-Leon, Morphological associative memories, IEEE Transactions on Neural Networks 9 (2) (1998) 281-293.

[40] G.X. Ritter and G. Urcid, Lattice algebra approach to single-neuron computation, IEEE Transactions on Neural Networks 14 (2) (2003) 282-295.

[41] G.X. Ritter, G. Urcid, and L. Iancu, Reconstruction of patterns from noisy inputs using morphological associative memories, Journal of Mathematical Imaging and Vision 19 (2) (2003) 95-111.

[42] P.K. Simpson, Fuzzy min-max neural networks - part1: classification. IEEE Transactions on Neural Networks 3(5) (1992) 776-786.

[43] P.K. Simpson, Fuzzy min-max neural networks - part2: clustering. IEEE Transactions on Fuzzy Systems 1(1) (1993) 32-45.

[44] P. Sussner, Observations on morphological associative memories and the kernel method, Neurocomputing 31 (1-4) (2000) 167-183.

[45] P. Sussner, Associative morphological memories based on variations of the kernel and dual kernel methods, Neural Networks 16 (5-6) (2003) 625-632.

[46] P. Sussner and M.E. Valle, Implicative fuzzy associative memories, IEEE Transactions on Fuzzy Systems 14 (6) (2006) 793-807.

[47] P. Sussner and M.E. Valle, Gray-scale morphological associative memories, IEEE Transactions on Neural Networks 17 (3) (2006) 559-570.

[48] P. Sussner and M.E. Valle, Morphological and certain fuzzy morphological associative memories for classification and prediction, in: V.G. Kaburlasos and G.X. Ritter, eds., Computational Intelligence Based on Lattice Theory (Springer, Heidelberg, Studies in Computational Intelligence 67, 2007) 149-171.

[49] K. Uehara and M. Fujise, Fuzzy inference based on families of α-level sets, IEEE Transactions on Fuzzy Systems 1 (2) (1993) 111-124.

[50] K. Uehara and K. Hirota, Parallel and multistage fuzzy inference based on families of α-level sets, Information Sciences 106 (1-2) (1998) 159-195.

[51] S.J. Verzi, G.L. Heileman, and M. Georgiopoulos, Boosted ARTMAP: modifications to fuzzy ARTMAP motivated by boosting theory, Neural Networks 19 (4) (2006) 446-468.

[52] Xiaodong Liu, Wei Wang, and Tianyou Chai, The fuzzy clustering analysis based on AFS theory, IEEE Transactions on Systems, Man and Cybernetics – Part B 35 (5) (2005) 1013-1027.

[53] Y. Xu, D. Ruan, K. Qin, and J. Liu, Lattice-Valued Logic (Springer-Verlag, Berlin, Studies in Fuzziness and Soft Computing 132, 2003).

[54] R.R. Yager, Toward a language for specifying summarizing statistics, IEEE Transactions on Systems, Man and Cybernetics – Part B 33 (2) (2003) 177-187.

[55] P.-F. Yang and P. Maragos, Min-max classifiers: learnability, design and application, Pattern Recognition 28 (6) (1995) 879-899.

[56] L.A. Zadeh, The concept of a linguistic truth variable and its application to approximate reasoning – I, II, III, Information Sciences 8 (3) (1975) 199-249; 8 (4) (1975) 301-357; 9 (1) (1975) 43-80.

[57] M. Zhong, B. Rosander, M. Georgiopoulos, G.C. Anagnostopoulos, M. Mollaghasemi, and S. Richie, Experiments with safe μARTMAP: effect of the network parameters on the network performance, Neural Networks 20 (2) (2007) 245-259.

Figure Captions

Fig. 1. Learning (i.e. training) by Fuzzy Lattice Reasoning (FLR).

Fig. 2. Generalization (i.e. testing) by Fuzzy Lattice Reasoning (FLR).

Fig. 3. (a) Sigmoid function $v_R(x) = 2/(1+e^{-3(x-4)})$.

(b) The derivative $6e^{-3(x-4)}/(1+e^{-3(x-4)})^2$ of the sigmoid function $v_R(x)$ above. A global maximum of 1.5 is attained at $x = 4$.

Fig. 4. Grayscale digital images of fertilizer granules on black background

(a) under white light, and

(b) under ultraviolet light.

Fig. 5. Optimally estimated sigmoid positive valuation functions in an industrial application

(a) $v_{R,1}(x)= 50.2839/(1+e^{-45.4816x})$ regarding input variable *area*.

(b) $v_{R,2}(x)= 0.7575/(1+e^{-0.9764x})$ regarding input variable *circularity*.

(c) $v_{R,3}(x)= 14.3893/(1+e^{-0.0038(x+110.5680)})$ regarding input variable *brightness difference*.

In addition, the following three rules are displayed:

Rule R1: IF A1.AND.C1.AND.BD1 THEN **small**.

Rule R2: IF A2.AND.C2.AND.BD2 THEN **medium**.

Rule R3: IF A3.AND.C3.AND.BD3 THEN **large**.

Inference is carried out by "winner-take-all" fuzzy lattice reasoning (FLR).

Table Captions

Table 1

In the binary case, where $A,B \in \{0,1\}$, the truth table of the implication function "$A \Rightarrow B$" is identical to the truth table of the inclusion measure function "$\sigma(A \leq B)$"

Table 2

Distribution of industrial data, regarding (fertilizer granule) *stiffness*, in three categories **small**/**medium**/**large** used for training/validation/testing

Table 3

A line displays a considered range (min/MAX values) as well as the optimally computed estimate for both sigmoid parameters A_i, λ_i, m_i in a data dimension i=1,2,3 and the overall critical granule size Z_{crit}

Table 4

Different classifiers, for various combinations of data types and positive valuation functions, have resulted in different numbers of misclassifications

Granular FLR, for training

FLR-0: A set of pairs $RB = \{(\boldsymbol{E}_1,c_1),...,(\boldsymbol{E}_L,c_L)\}$ is given, where $\boldsymbol{E}_l \in \mathsf{F}^N$ is a granule/cluster and $c_l \in \mathsf{C}$, $l=1,...,L$ is its corresponding class label.

FLR-1: Do "set" all pairs in *RB*. Present the next input pair $(\boldsymbol{F}_i,c_i) \in \mathsf{F}^N \times \mathsf{C}$, $i=1,...,n$. Compute the degree of inclusion $\sigma(\boldsymbol{F}_i \leq \boldsymbol{E}_l)$ of input granule $\boldsymbol{F}_i$ to all granules $\boldsymbol{E}_l$, $l=1,...,L$.

FLR-2: If no more pairs are "set" in *RB* then store input pair $(\boldsymbol{F}_i,c_i)$ in *RB*; $L \leftarrow L+1$; goto step FLR-1.

FLR-3: Competition among the "set" pairs in *RB*: Winner is pair $(\boldsymbol{E}_J,c_J)$ such that $J \doteq \arg \max_{l \in \{1,...,L\}} \sigma(\boldsymbol{F}_i \leq \boldsymbol{E}_l)$. In case of multiple winners, choose the one with the smallest (granule) size.

FLR-4: The *Assimilation Condition*: Both 1) the size $Z(\boldsymbol{F}_i \vee \boldsymbol{E}_J)$ of granule $\boldsymbol{F}_i \vee \boldsymbol{E}_J$ is less than a user-defined threshold size Z_{crit}, and 2) $c_i = c_J$.

FLR-5: If the *Assimilation Condition* is not satisfied then "reset" the winner pair $(\boldsymbol{E}_J,c_J)$; goto step FLR-2.

Else, replace the winner granule $\boldsymbol{E}_J$ with granule $\boldsymbol{F}_i \vee \boldsymbol{E}_J$; goto step FLR-1.

Fig. 1.

Granular FLR, for testing

flr-0: Consider a set of pairs $RB = \{(\boldsymbol{E}_1, c_1), ..., (\boldsymbol{E}_L, c_L)\}$, where $\boldsymbol{E}_l \in \mathsf{F}^N$ is a granule/cluster and $c_l \in \mathsf{C}$, $l=1,...,L$ is its corresponding class label.

flr-1: Present a granule $\boldsymbol{F}_0 \in \mathsf{F}^N$.

flr-2: Compute the degree of inclusion $\sigma(\boldsymbol{F}_0 \leq \boldsymbol{E}_l)$ of granule $\boldsymbol{F}_0$ to all granules $\boldsymbol{E}_l$, $l=1,...,L$.

flr-3: Competition among the pairs in RB: Winner is pair $(\boldsymbol{E}_J, c_J)$ such that $J \doteq \arg \max_{l \in \{1,...,L\}} \sigma(\boldsymbol{F}_0 \leq \boldsymbol{E}_l)$. In case of multiple winners, choose the one with the smallest (granule) size.

flr-4: Granule $\boldsymbol{F}_0$ is classified to the class with label c_J.

Fig. 2.

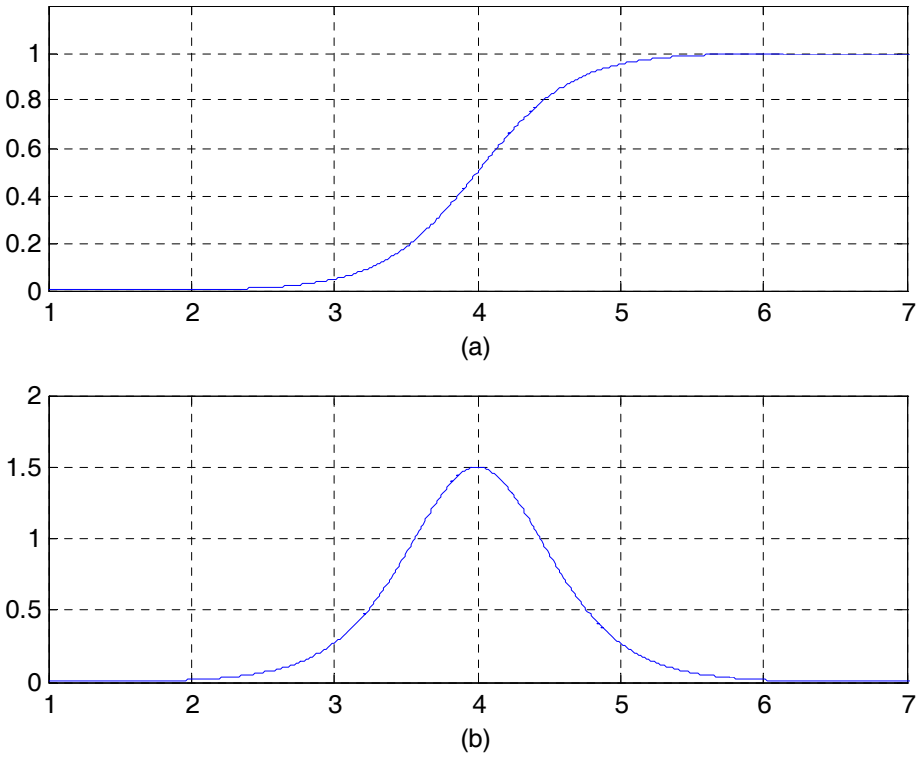
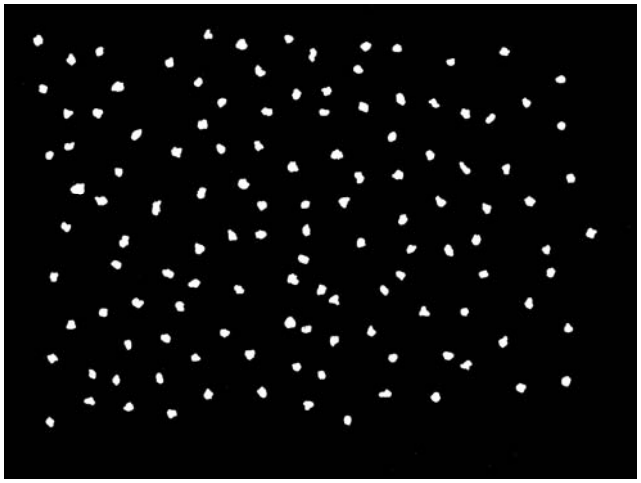
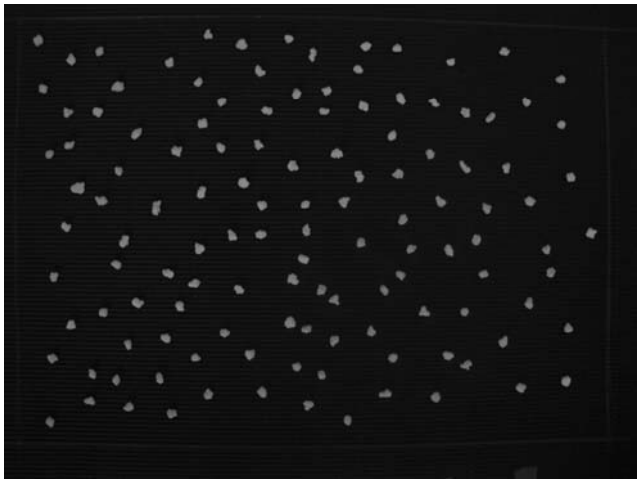


Fig. 3.



(a)



(b)

Fig. 4.

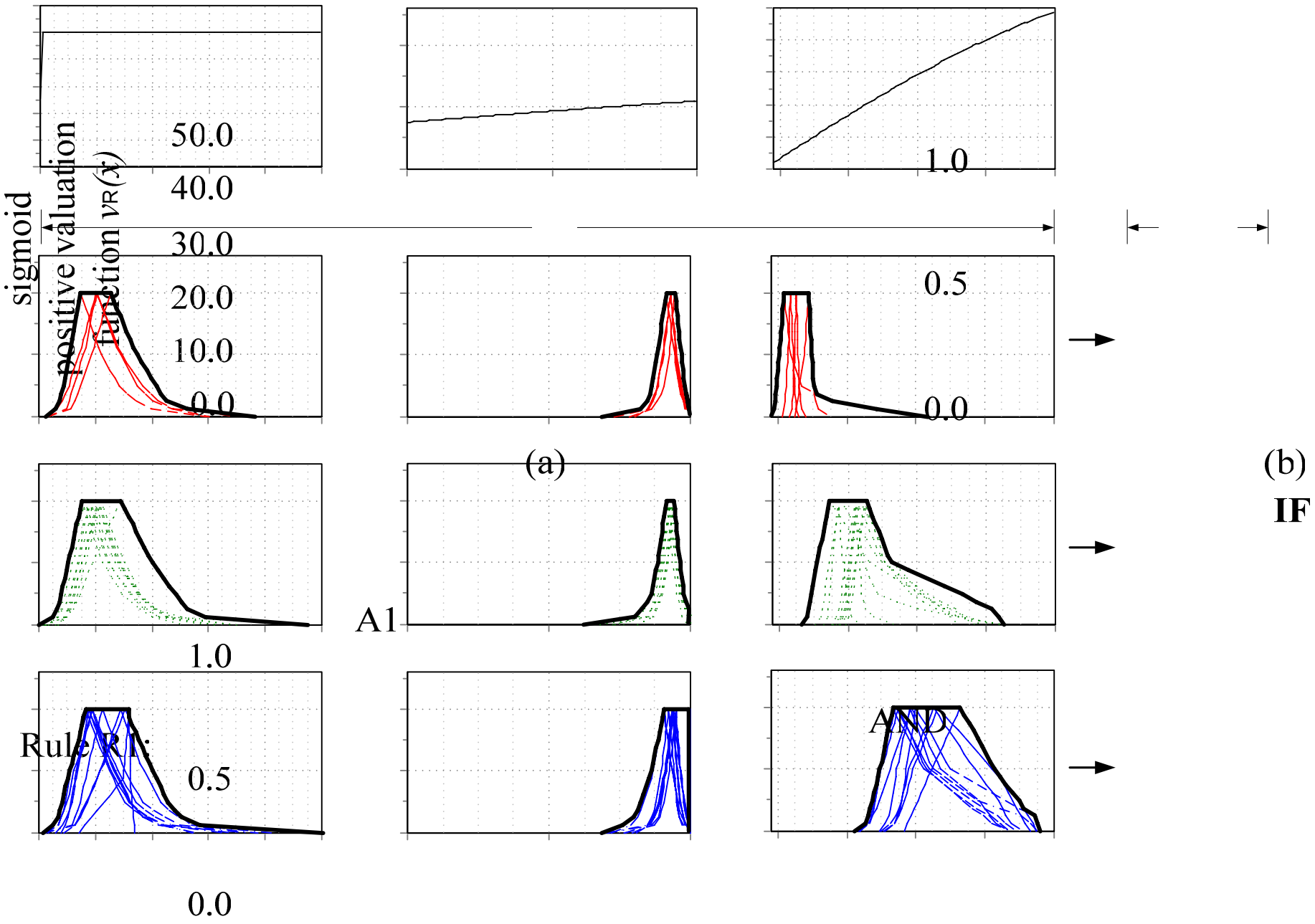

sigmoid
positive valuation
function $\nu_R(x)$
50.0
40.0
30.0
20.0
10.0
0.0
1.0
0.5
0.0
(a)
(b)
IF
A1
A2
AND
Rule R1:
1.0
0.5
0.0

Fig. 5.

Table 1

In the binary case, where $A,B \in \{0,1\}$, the truth table of the implication function "$A \Rightarrow B$" is identical to the truth table of the inclusion measure function "$\sigma(A \leq B)$"

A	B	$A \Rightarrow B$	$\sigma(A \leq B)$
0	0	1	1
0	1	1	1
1	0	0	0
1	1	1	1

Table 2

Distribution of industrial data, regarding (fertilizer granule) *stiffness*, in three categories **small**/**medium**/**large** used for training/validation/testing

	categories of (fertilizer granule) *stiffness*			TOTAL
no. of data used for	**small**	**medium**	**large**	no. of data
training	5	10	10	25
validation	5	3	2	10
testing	1	4	2	7
TOTAL no. of data	11	17	14	42

Table 3

A line displays a considered range (min/MAX values) as well as the optimally computed estimate for both sigmoid parameters A_i, λ_i, m_i in a data dimension i=1,2,3 and the overall critical granule size Z_{crit}

parameter	parameter range		optimal parameter
	min value	MAX value	estimate
A_1	0	100	50.2839
λ_1	0	100	45.4816
m_1	0	100	0.0
A_2	0	2	0.7575
λ_2	0	2	0.9764
m_2	0	2	0.0
A_3	0	600	14.3893
λ_3	0	600	0.0038
m_3	-120	600	-110.5680
Z_{crit}	0	10	5.0

Table 4

Different classifiers, for various combinations of data types and positive valuation functions, have resulted in different numbers of misclassifications

classifier name	data type	positive valuation function $v(x)$	no. of misclassifications
FLR	FIN	sigmoid	0
FLR	FIN	$v(x) = x$	1
kNN	FIN	sigmoid	1
kNN	FIN	$v(x) = x$	1
kNN	number	sigmoid	1
kNN	number	$v(x) = x$	1