Piecewise-Linear Approximation of Nonlinear Models Based on Interval Numbers (INs)

Vassilis G. Kaburlasos and S. E. Papadakis

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

Abstract. Linear models are ususally preferable due to their simplicity. However, nonlinear models often emerge in practice. A popular approach for dealing with nonlinearities is using a piecewise-linear approximation. In such context, inspired from both Fuzzy Inference Systems (FISs) of TSK type and Self-Organizing Maps (SOMs), this work introduces enhancements based on Interval Numbers and, ultimately, on lattice theory. Advantages include a capacity to deal with *granular* inputs, introduction of *tunable* nonlinearities, representation of *all-order* statistics, and induction of descriptive decision-making knowledge (*rules*) from the training data. Preliminary computational experiments here demonstrate a good capacity for generalization; furthermore, only a few rules are induced.

Key words: Fuzzy inference systems (FIS), Genetic optimization, Granular data, Interval number (IN), Lattice theory, Linear approximation, Rules, Self-organizing map (SOM), TSK model

1 Introduction

The need to employ a real *function* $y : \mathsf{R}^N \rightarrow \mathsf{R}^M$, i.e. a *model*, arises frequently in practice. In particular, linear models $y(\mathbf{x}) = c_0 + c_1x_1 + c_2x_2 + ... + c_Nx_N$ are preferable due to simplicity. However, most often, the dependence of output y on the input variables $x_1, ..., x_N$ is nonlinear.

A popular approach for dealing with nonlinearities is using a piecewise-linear approximation. For instance, in the context of fuzzy logic, the TSK (Tagaki-Sugeno-Kang) fuzzy model [13], [14], [15], [16] is popular. The computation of a TSK model, in the first place, involves the computation of clusters.

A popular scheme for clustering is the *self-organizing map* (SOM) devised for visualization of nonlinear relations of multidimensional data [10]. Lately, *granular* extensions of SOM were proposed in classification applications [8], [11], where a data cluster was represented by a *fuzzy interval number* (FIN).

This work proposes simpler acronym IN (Interval Number) for a FIN. In the sequel, it explains that a IN is a mathematical object, which may be interpreted as a probability/possibility distribution, an interval, and/or a real number. In conclusion, inspired from TSK modeling, this work proposes *lattice computing*

pp. 13–22, ISBN 978–80–244–2112–4, Palacký University, Olomouc, 2008.

techniques for an advantageous, piecewise-linear approximation based on a IN-extension of SOM.

We remark that the term *lattice computing*, or *LC* for short, was coined lately to denote an emerging Computational Intelligence paradigm based on lattice theory [3]. More accurately, LC is defined here 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.

This paper is organized as follows. Section 2 describes the mathematical background. Section 3 outlines the proposed techniques. Section 4 presents preliminary experimental results. Section 5 concludes by summarizing the contribution. The Appendix summarizes the WRLS algorithm for incremental learning.

2 Mathematical Background

Here we summarize useful mathematical notions and tools regarding Interval Numbers (INs) [4], [6], [7], [8] using an improved mathematical notation [9].

2.1 The Vector Lattice $(\mathcal{\Delta},\leq)$ of Generalized Intervals

Assume the *complete latice* $(\mathsf{R},\leq)$ of real numbers with *least* and *greatest* elements denoted, respectively, by $O=-\infty$ and $I=+\infty$. A *generalized interval* is defined in the following.

Definition 1. *A* generalized interval *is an element of lattice* $(\mathsf{R},\leq^{\partial})\times(\mathsf{R},\leq)$.

We remark that $\leq^{\partial}$ in Definition 1 denotes the dual (i.e. converse) of order relation $\leq$, i.e. $\leq^{\partial}\equiv\geq$. Moreover, product lattice $(\mathsf{R},\leq^{\partial})\times(\mathsf{R},\leq)\equiv(\mathsf{R}\times\mathsf{R},\geq\times\leq)$ will be denoted by $(\mathcal{\Delta},\leq)$.

A generalized interval is denoted by $[x,y]$, where $x,y\in\mathsf{R}$. Apparently, the corresponding *meet* and *join* in lattice $(\mathcal{\Delta},\leq)$ are given, respectively, by $[a,b]\wedge[c,d]=[a\vee c,b\wedge d]$ and $[a,b]\vee[c,d]=[a\wedge c,b\vee d]$, where $a\wedge c$ $(a\vee c)$ denotes the *minimum* (*maximum*) of real numbers a and c.

The set of *positive* (*negative*) generalized intervals $[a,b]$, characterized by $a\leq b$ $(a>b)$, is denoted by $\mathcal{\Delta}_{+}$ $(\mathcal{\Delta}_{-})$. Apparently, lattice $(\mathcal{\Delta}_{+},\leq)$ of *positive* generalized intervals is *isomorphic*[1] to the lattice $(\tau(\mathsf{R}),\leq)$ of intervals (sets) in R, i.e. $(\tau(\mathsf{R}),\leq)\cong(\mathcal{\Delta}_{+},\leq)$. We have augmented lattice $(\tau(\mathsf{R}),\leq)$ by a *least* (empty) interval, denoted by $O=[+\infty,-\infty]$. Note that a *greatest* interval $I=[-\infty,+\infty]$

[1] A map $\psi:(\mathsf{P},\leq)\rightarrow(\mathsf{Q},\leq)$ is called *(order) isomorphism* if and only if both "$x\leq y\Leftrightarrow\psi(x)\leq\psi(y)$" and "$\psi$ is onto Q". Two lattices $(\mathsf{P},\leq)$ and $(\mathsf{Q},\leq)$ are called *isomorphic*, symbolically $(\mathsf{P},\leq)\cong(\mathsf{Q},\leq)$, if and only if there is an isomorphism between them.

already exists in $\tau(\mathsf{R})$. Hence, the complete lattice $(\tau_O(\mathsf{R}) = \tau(\mathsf{R}) \cup \{O\},\leq)$ emerges — For simplicity, we use symbols O and I to denote the least and greatest element, respectively, in any complete lattice.

A (strictly) decreasing *bijective*, the 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. $x < y \Leftrightarrow \theta_{\mathsf{R}}(x) > \theta_{\mathsf{R}}(y)$, $x, y \in \mathsf{R}$. Furthermore, a strictly increasing function $v_{\mathsf{R}} : \mathsf{R} \to \mathsf{R}$ is a *positive valuation*[2] 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 $(\Delta,\leq)$ [5]. It follows a metric function $d_{\Delta} : \mathsf{R} \to \mathsf{R}^{\geq 0}$ given by $d_{\Delta}([a,b],[c,d]) = v_{\Delta}([a,b] \vee [c,d]) - v_{\Delta}([a,b] \wedge [c,d]) = [v_{\mathsf{R}}(\theta_{\mathsf{R}}(a \wedge c)) - v_{\mathsf{R}}(\theta_{\mathsf{R}}(a \vee c))] + [v_{\mathsf{R}}(b \vee d) - v_{\mathsf{R}}(b \wedge d)]$. In particular, metric d_{Δ} is valid in lattice $(\Delta_+ \cup \{O\},\leq) \cong (\tau_O(\mathsf{R}),\leq)$.

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 $v_{\Delta}([a,b]) = v_{\mathsf{R}}(b) - v_{\mathsf{R}}(a)$; hence, it follows metric $d_{\Delta}([a,b],[c,d]) = [v_{\mathsf{R}}(a \vee c) - v_{\mathsf{R}}(a \wedge c)] + [v_{\mathsf{R}}(b \vee d) - v_{\mathsf{R}}(b \wedge d)]$ [6]. In particular, for $\theta_{\mathsf{R}}(x) = -x$ and $v_{\mathsf{R}}(x) = x$ it follows metric $d_{\Delta}([a,b],[c,d]) = |a-c| + |b-d|$. In general, *parametric* functions $\theta_{\mathsf{R}}(.)$ and $v_{\mathsf{R}}(.)$ may imply tunable nonlinearities.

The space Δ of generalized intervals is a *real linear space* [4], [8] with

- *addition* defined as $[a,b] + [c,d] = [a+c, b+d]$.
- *multiplication* (by a scalar $k \in \mathsf{R}$) defined as $k[a,b] = [ka, kb]$.

A generalized interval in real linear space Δ is also called *vector*. A lattice-ordered vector space is called *vector lattice* [4].

A subset C of a linear space is called *cone* if and only if for $x_1, x_2 \in C$ and real numbers $\lambda_1, \lambda_2 \geq 0$ it follows $(\lambda_1 x_1 + \lambda_2 x_2) \in C$. It turns out that set Δ_+ is a cone. Likewise, set Δ_- is a cone.

2.2 The Cone Lattice $(\mathsf{F},\leq)$ of Interval Numbers (INs)

Generalized interval analysis in the previous section is useful for studying *interval numbers* (INs) in this section. A more general number type is defined first, in the following.

Definition 2. *A* generalized interval number*, or* GIN *for short, is a function* $G : (0,1] \to \Delta$.

Let G denote the set of GINs. It turns out that $(\mathsf{G},\leq)$ is a complete lattice since $(\mathsf{G},\leq)$ is the Cartesian product of complete lattices $(\Delta,\leq)$.

Addition and multiplication are extended from Δ to G as follows.

- The *sum* $G_1 + G_2$, $G_1, G_2 \in \mathsf{G}$ is defined as $G_s : G_s(h) = (G_1 + G_2)(h) = G_1(h) + G_2(h)$, $h \in (0,1]$.

[2] *Positive valuation* is a function $v : (\mathsf{L},\leq) \times (\mathsf{L},\leq) \to \mathsf{R}$, which satisfies both $v(x) + v(y) = v(x \wedge y) + v(x \vee y)$ and $x < y \Rightarrow v(x) < v(y)$.

– The *product* kG_1, $k \in \mathsf{R}$ and $G_1 \in \mathsf{G}$, is defined as $G_p : G_p(h) = kG_1(h)$, $h \in (0,1]$.

Our interest here focuses on the *sublattice*[3] of *interval numbers* defined next.

Definition 3. *An* Interval Number*, or* IN *for short, is a GIN F such that both $F(h) \in (\Delta_+ \cup \{O\})$ and $h_1 \leq h_2 \Rightarrow F(h_1) \geq F(h_2)$.*

Let F denote the set of INs. Conventionally, a IN will be denoted by a capital letter in italics, e.g. $F \in \mathsf{F}$. Moreover, a N-tuple IN will be denoted by a capital letter in bold, e.g. $\mathbf{F} = (F_1, ..., F_N) \in \mathsf{F}^N$.

From Definition 3 it follows that a general IN F is written as the set union of (conventional) intervals, e.g. $F = \underset{h\in(0,1]}{\cup}\{[a_h, b_h]\}$, where both interval-ends a_h and b_h are functions of $h \in (0,1]$ such that $a_h \leq b_h$.

We point out that a IN is a mathematical object, which may be interpreted as a probability/possibility distribution, an interval, and/or a real number. For instance, IN $F = \underset{h\in(0,1]}{\cup}\{[a,b]\}$ represents interval $[a,b]$ including real numbers for $a = b$. Moreover, a IN F may represent a probability distribution such that interval $F(h)$ includes $100(1-h)\%$ of the distribution, whereas the remaining $100h\%$ is split even both below and above interval $F(h)$ [4], [7], [8]. In addition, a IN may represent a fuzzy number as explained in subsection 2.3 below. In all cases, a IN can be interpreted as a *granule* (of information).

It has been shown that for $F, E \in \mathsf{F}$ there follow both $(F \wedge E) \in \mathsf{F}$ and $(F \vee E) \in \mathsf{F}$ [9]. Hence, $(\mathsf{F}, \leq)$ is a lattice with ordering $F_1 \leq F_2 \Leftrightarrow F_1(h) \leq F_2(h), \forall h \in (0,1]$.

The following proposition introduces a metric in lattice $(\mathsf{F}, \leq)$ based on a positive valuation function $v_{\mathsf{R}} : \mathsf{R} \to \mathsf{R}^{\geq 0}$ [9].

Proposition 1. *Let F_1 and F_2 be INs in the lattice $(\mathsf{F}, \leq)$ of INs. Assuming that the following integral exists, a metric function $d_{\mathsf{F}} : \mathsf{F} \times \mathsf{F} \to \mathsf{R}^{\geq 0}$ is given by*

$$d_{\mathsf{F}}(F_1, F_2) = \int_0^1 d_{\Delta}(F_1(h), F_2(h))dh \tag{1}$$

We remark that a Minkowski metric $d_p : \mathsf{F}^N \times \mathsf{F}^N \to \mathsf{R}^{\geq 0}$ can be defined between two N-tuple INs $\mathbf{F}_1 = [F_{1,1}, ..., F_{1,N}]^T$ and $\mathbf{F}_2 = [F_{2,1}, ..., F_{2,N}]^T$ as

$$d_p(\mathbf{F}_1, \mathbf{F}_2) = [d^p_{\mathsf{F}}(F_{1,1}, F_{2,1}) + ... + d^p_{\mathsf{F}}(F_{1,N}, F_{2,N})]^{1/p} \tag{2}$$

Minkowski metric $d_p(\mathbf{F}_1, \mathbf{F}_2)$ may involve a point $\mathbf{x} = [x_1, ..., x_N]^T \in \mathsf{R}^N$ such that an entry x_i is represented by *trivial* IN $x_i = \underset{h\in(0,1]}{\cup}\{[x_i, x_i]\}$, $i = 1, ..., N$.

Space F is a *cone* for $F_1, F_2 \in \mathsf{F}$ and real numbers $\lambda_1, \lambda_2 \geq 0$ it follows $(\lambda_1 F_1 + \lambda_2 F_2) \in \mathsf{F}$.

[3] A *sublattice* of a lattice $(\mathsf{L}, \leq)$ is another lattice $(\mathsf{S}, \leq)$ such that $\mathsf{S} \subseteq \mathsf{L}$.

2.3 Perspectives

A fundamental result in *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 [19]. The aforementioned theorem was given little attention in practice, to-date. However, some authors have capitalized on it by designing fuzzy inference systems (FIS) based on α-cuts of fuzzy numbers, i.e. based on intervals in $\tau(\mathsf{R})$ [17], [18]. 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 by, first, dropping the possibilistic interpretation for a (fuzzy) membership function and, second, by considering its equivalent α-cuts (interval) representation.

3 The Proposed Techniques

This section outlines computational techniques for achieving a piecewise-linear approximation of nonlinear models based on INs. Further details will be presented in a future publication.

3.1 Structure Identification

Structure identification is a term from "fuzzy TSK system modeling" [12],[15],[16] meaning a partition of a model's input space in *subspaces*, or *clusters*, such that the output to an "input point $\mathbf{x} = [x_1, ..., x_N]^T$, within a cluster" is a (usually) linear combination of the N inputs $x_1, ..., x_N$. It turns out that the task of *structure identification* is not trivial as illustrated in the following.

Consider the data points shown together with piecewise-linear approximations of two different single-input-single-output models in Fig. 1(a) and Fig. 1(b), respectively. Fig. 1(a) demonstrates an *effective* partition (of the input space) characterized by a small approximation error, whereas Fig. 1(b) demonstrates an *ineffective* partition characterized by a large approximation error.

A structure identification method is proposed next based (1) on a novel SOM extension, and (2) on an advantageous, novel structure identification algorithm.

3.2 A SOM Extension

Each cell $C_{i,j}$ in the SOM proposed here stores both a N-dimensional IN $\mathbf{F}_{i,j} = [F_{i,j,1}, ..., F_{i,j,N}]^T$ and a $(N+1)$-dimensional vector $\mathbf{c}_{i,j} = [c_{i,j,0}, c_{i,j,1}, ..., c_{i,j,N}]^T$, where $i = 1, ..., I$, and $j = 1, ..., J$. On one hand, IN $\mathbf{F}_{i,j} \in \mathsf{F}^N$ represents a population of data assigned to cell $C_{i,j}$. On the other hand, vector $\mathbf{c}_{i,j} \in \mathsf{R}^{N+1}$ stores the parameters of the following hyperplane

$$p_{i,j}(\mathbf{x}) = c_{i,j,0} + c_{i,j,1}x_1 + c_{i,j,2}x_2 + ... + c_{i,j,N}x_N \tag{3}$$

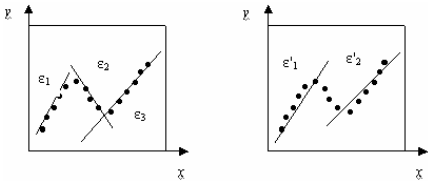

Fig. 1. Two different, piecewise-linear single-input-single-output models. (a) This model partitions the input space *effectively* with a small approximation error using three lines. (b) This model partitions the input space *ineffectively* with a large approximation error using two lines.

A cell is called *nonempty* if at least one datum is assigned to it. A nonempty cell represents a rule. In particular, the N INs in $\mathbf{F}_{i,j}$ of cell $C_{i,j}$ correspond to a (fuzzy) rule antecedent, whereas the $N+1$ hyperplane parameters in $\mathbf{c}_{i,j}$ constitute the corresponding rule's consequent.

3.3 INSOM: A Structure Identification Algorithm

Structure identification is carried out using the novel algorithm INSOM, below.

3.4 Parameter Identification

Algorithm INSOM above induces an "initial" (piecewise-linear) model from a series $(\mathbf{x}_k, y_k) \in \mathsf{R}^N \times \mathsf{R}$, $k = 1, 2, ..., n$ of training data. The objective in this section is to compute a globally optimum model.

The output of the aforementioned "initial" model is written analytically as

$$\widehat{y}(\mathbf{x}_k) = c_0 + \sum_{i=1}^{L} (c_{i,0}\sigma_i + \sum_{j=1}^{N} c_{i,j}\sigma_i x_{k,j}) \tag{4}$$

where $\mathbf{x}_k = [x_{k,1}, ..., x_{k,N}]^T$, furthermore the σ_is are functions of the (known) INs. In conclusion, a globally optimum set of hyperplanes is computed by algorithm WRLS in the Appendix.

Further improvement was sought by optimal parameter estimation techniques, which replaced a IN $F_{i,j}$ by IN $F'_{i,j} = a_{i,j}F_{i,j} + b_{i,j}$, where $a_{i,j} \in (0, 3]$ is a *scaling* parameter and $b_{i,j} \in [-1, 1]$ is a *translation* parameter, $i = 1, ..., L, j = 1, ..., N$. More specifically, the task was to compute optimal INs $F'_{i,j}$, in a mean square error sense, by optimal parameter $a_{i,j}$, $b_{i,j}$ estimation.

Optimization was pursued by genetic algorithms (GA) [1],[12], where the phenotype of an "individual" consisted of specific values of parameters $a_{i,j}, b_{i,j}$. There was a total number of $2 \times N \times L$ parameters binary-encoded to the chromosome of an "individual". We included 25 "individuals" per generation.

In conclusion, we point out that our "initial" model was computed by algorithm INSOM for structure identification without any employment of fuzzy

Algorithm 1 INSOM: A Structure Identification Algorithm

$I \leftarrow$ Number of rows in a SOM grid/map
$J \leftarrow$ Number of columns in a SOM grid/map
$d_\theta \leftarrow 0.8$, $\ell_\theta \leftarrow n/10$ *//user-defined parameters d_θ and ℓ_θ*
createANDinitializeMap(I, J)
for $r = 1$ to N_{epochs} **do** *//for each epoch*
  Calculate $B_{p,q}(r), a(r)$ *//$B_{p,q}(r)$ is a neighborhood; $a(r)$ is a weight coefficient*
  $w_k \leftarrow a(r)$
  for $k = 1$ to n **do** *//for each input datum $(\mathbf{x}_k, y_k)$*
    FindTheWinner$(\mathbf{x}_k, y_k)$
    $p \leftarrow$ winner row
    $q \leftarrow$ winner column
    Assign$(\mathbf{x}_k, y_k, p, q)$ *//assign input datum $(\mathbf{x}_k, y_k)$ to winner cell $C_{p,q}$*
    for $i = 1$ to I **do** *//for each row*
      for $j = 1$ to J **do** *//for each column*
        if $C_{i,j} \in B_{p,q}(r)$ **then**
          WRLS$(i, j, w_k, \mathbf{x}_k, y_k)$
        end if
      end for//*for j*
    end for//*for i*
  end for//*for k*
  ResetCellsConditionally(ℓ_θ)
  ComputeINs()
  MergeSimilarCells(d_θ)
end for//*for r*

logic. Whereas, thereafter, parameter identification was pursued based on standard fuzzy TSK modeling techniques.

4 Preliminary Experimental Results

The effectiveness of our proposed (piecewise-linear approximation) techniques is demonstrated in this preliminary work on a single-input-single-output nonlinear system. In the interest of simplicity positive valuation function $v_{\mathsf{R}}(x) = x$ was employed. Furthermore, both input- and output- data were normalized in the interval $[0, 1]$ by straightforward linear transformation. At the end of all computations, the output data were transformed back to their original domain for meaningful comparisons.

We considered the simple system described by the following equation.

$$y = sin(10x) \tag{5}$$

where $x \in [0, 1]$.

Forty input/output data pairs $(x_k, y_k) \in \mathsf{R} \times \mathsf{R}$, $k = 1, ..., 40$ were randomly (uniformly) generated. The scatter plot of the generated input/output data points is shown in Fig. 2(a). Following a popular practice, we employed the same data set for both training and testing. No validation set was employed.

A 4×4 SOM grid was used to compute a TSK model. The structure identification algorithm was applied for $N_{epochs} = 100$ epochs resulting in five nonempty cells — Recall that a nonempty cell represents a rule. The IN/antecedent and the hyperplane/consequent (the latter is a line here) in each cell are shown in Fig. 2(b) and Fig. 2(a), respectively. A visual inspection of Fig. 2 reveals that the proposed method partitions the input space well.

5 Conclusion

This work has proposed a new paradigm, inspired from both Fuzzy Inference Systems (FISs) of TSK type and Self-Organizing Maps (SOMs), for piecewise-linear approximation of nonlinear models based on Interval Numbers (INs).

A unique advantage of INs here is their effectiveness in computing colinear points within a cluster as it will be detailed in a future publication. Another advantage of our proposed techniques is the fast induction of an optimal number of rules. Note that the employment of SOM in fuzzy system modeling applications has been rather sporadic to-date. Nevertheless, different authors have confirmed the capacity of SOM for rapid data processing [2]. In our future work we have also planned additional experiments including alternative data sets.

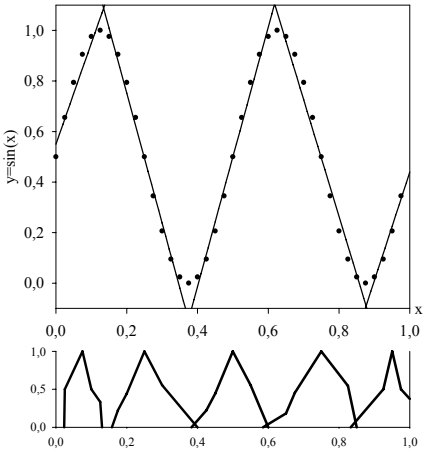

Fig. 2. (a) Scatter plot of function $y = sin(10x)$ including 40 input/output data points. The five lines correspond, respectively, to the consequents of five rules. (b) The five INs correspond, respectively, to the antecedents of five rules — Note that the corresponding consequent (line) for a IN is shown above the IN.

Appendix

Here we show the Weighted Recursive Least Squares (WRLS) algorithm for incremental learning.

Consider a series of data vectors $[x_{k,1}, ..., x_{k,M}, y_k]^T \in \mathsf{R}^M \times \mathsf{R}$, $k = 1, ..., n$. The WRLS algorithm computes incrementally the parameters of a hyperplane in R^{M+1}, *optimally fitted*, in a least square error sense, to the aforementioned data. The corresponding equations are shown next.

$$\begin{aligned} \mathbf{c}_{k+1} &= \mathbf{c}_k + \left(y_{k+1} - \mathbf{x}_{k+1}^T \cdot \mathbf{c}_k\right) \mathbf{k}_k \\ \mathbf{k}_k &= \frac{\mathbf{S}_k \mathbf{x}_{k+1}}{\frac{1}{w_k} + \mathbf{x}_{k+1}^T \mathbf{S}_k \mathbf{x}_{k+1}} \\ \mathbf{S}_{k+1} &= \left(\mathbf{I} - \mathbf{k}_k \mathbf{x}_{k+1}^T\right) \mathbf{S}_k \\ k &= 1, 2, ..., n. \end{aligned} \tag{6}$$

The equations above are initialized at $k = 0$ with $\mathbf{c}_0 = \mathbf{0}$ and $\mathbf{S}_0 = a\mathbf{I}$, where $a \in \mathsf{R}$ is typically large, e.g. $a = 1000$. Vector $\mathbf{c}_k = [c_{k,0}, c_{k,1}, ..., c_{k,M}]^T$ includes the *optimum* hyperplane parameters at a step.

References

1. Cordón, O., Gomide, F., Herrera, F., Hoffmann, F., Magdalena, L.: Ten years of genetic fuzzy systems: current framework and new trends. Fuzzy Sets and Systems **141(1)** (2004) 5–31

2. Er, M. J., Li, Z., Cai, H., Chen, Q.: Adaptive noise cancellation using enhanced dynamic fuzzy neural networks. IEEE Trans. Fuzzy Systems **13(3)** (2005) 331–342
3. Graña, M.: Lattice computing: lattice theory based computational intelligence. In: T. Matsuhisa, H. Koibuchi (eds) Proc. Kosen Workshop on Mathematics, Technology, and Education (MTE), Irabaki National College of Technology, Ibaraki, Japan, 15-18 February 2008, pp. 19–27
4. Kaburlasos, V. G.: Towards a Unified Modeling and Knowledge-Representation Based on Lattice Theory. Springer, Heidelberg, Germany, ser Studies in Computational Intelligence **27** (2006)
5. Kaburlasos, V. G., Athanasiadis, I. N., Mitkas, P. A.: Fuzzy lattice reasoning (FLR) classifier and its application for ambient ozone estimation. Intl. J. Approximate Reasoning **45(1)** (2007) 152–188
6. Kaburlasos, V. G., Kehagias, A.: Novel fuzzy inference system (FIS) analysis and design based on lattice theory. part I: working principles. Intl. J. General Systems **35(1)** (2006) 45–67
7. Kaburlasos, V. G., Kehagias, A.: Novel fuzzy inference system (FIS) analysis and design based on lattice theory. IEEE Trans. Fuzzy Systems **15(2)** (2007) 243–260
8. Kaburlasos, V. G., Papadakis, S. E.: Granular self-organizing map (grSOM) for structure identification. Neural Networks **19(5)** (2006) 623–643
9. Kaburlasos, V. G., Papadakis, S. E.: A granular extension of the fuzzy-ARTMAP (FAM) neural classifier based on fuzzy lattice reasoning (FLR). Neurocomputing (accepted).
10. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg, Germany, ser Information Sciences **30** (1995)
11. Papadakis, S. E., Kaburlasos, V. G.: Mass-grSOM: a flexible rule extraction for classification. In: Proc. 5th Workshop Self-Organizing Maps (WSOM 2005), Paris, France, 5-8 September 2005, pp. 553–560
12. Papadakis, S. E., Theocharis, J. B.: A GA-based fuzzy modeling approach for generating TSK models. Fuzzy Sets and Systems **131(2)** (2002) 121–152
13. Sugeno, M., Kang, G. T.: Fuzzy modelling and control of multilayer incinerator. Fuzzy Sets Systems **18(3)** (1986) 329–346
14. Sugeno, M., Tanaka, K.: Successive identification of a fuzzy model and its applications to prediction of a complex system. Fuzzy Sets Systems **42(3)** (1991) 315–334
15. Sugeno, M., Yasukawa, T.: A fuzzy-logic-based approach to qualitative modeling. IEEE Trans. Fuzzy Systems **1(1)** (1993) 7–31
16. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Systems, Man & Cybernetics **15(1)** (1985) 116–132
17. Uehara, K., Fujise, M.: Fuzzy inference based on families of α-level sets. IEEE Trans. Fuzzy Systems **1(2)** (1993) 111–124
18. Uehara, K., Hirota, K.: Parallel and multistage fuzzy inference based on families of α-level sets. Information Sciences **106(1-2)** (1998) 159–195
19. Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning - III. Information Sciences **9(1)** (1975) 43–80