

Evolving Ensemble of Fuzzy Models

Eng Yeow Cheu[†], Chai Quek^{*} and See Siong Ng[†]

^{*}Centre for Computational Intelligence

Nanyang Technological University, Singapore

Email: ashcquek@ntu.edu.sg

[†]Institute for Infocomm Research

A*STAR (Agency for Science, Technology and Research), Singapore

Email: engyeow_cheu@pmail.ntu.edu.sg, skng@i2r.a-star.edu.sg

Abstract—This paper presents an online learning-based neuro-fuzzy system called evolving Fuzzy Ensemble (eFE). The hierarchical computational structure of eFE is progressively adapted to autonomously support fuzzy data associations in accordance with neurophysiological studies. Activity-dependent synapse with global decay learning rule is incorporated to simulate the retention and active forgetting mechanisms that are involved in memory persistence. Such features incorporated in eFE model make it suitable to address the nonstationary characteristics of real-world problems. This work demonstrates the use of simple mechanisms to accomplish complex form of associative learning, an idea that has been suggested by psychologists for many years but has only recently been verified at the cellular level. The proposed eFE model is evaluated and compared with other modelling techniques in two benchmark time series experiments. The experimental results demonstrate the capabilities, and illustrate the viability of the proposed modelling technique.

Index Terms—Fuzzy system, neuro-fuzzy system, ensemble, associative learning, online learning.

I. INTRODUCTION

Soft computing has been a significant and influential computing paradigm in drawing complementary methodologies to exploit the tolerance for uncertainty, so as to achieve cost-effective intelligent solutions [1]. A typical approach is the integration of neural network and fuzzy system to create a hybrid structure called *neuro-fuzzy system* (NFS) [2]. The application of NFS has been demonstrated in a number of fields, including time-varying signal detection [3], adaptive nonlinear control [4], forecasting [5], knowledge extraction [6], decision support [7], fault detection [8], and robotics [9]. This neural-fuzzy synthesis provides flexible information processing capability, in a human comprehensible neural construct, for handling real life complex and uncertain situations [2].

A. Incremental Adaptation

Existing NFSs (see, e.g., [10], [5], [11], [12], for instance) generally employ all the training data samples multiple times either during network structure identification or network parameter adaptation. These systems generally require the representative data samples of the problem domain to be collected in advance. Hence, such batch learning models are not suitable for modelling of complex time-varying processes in dynamically changing environments. In addition, the ever increasing flow of data makes batch learning models unfeasible because these models generally make use of time consuming

complex optimization with the objective of deriving a globally optimized solution that is representative of the training data. In order to use these models in online applications, it is only possible for the price of re-training the whole model structure and parameters with iterative and time-consuming procedures [13], [14].

In fact, a situation is likely to be encountered where it is not easy to collect a complete set of training samples to build a classification or forecasting system. For instance, it is almost impossible to construct a human face database without taking into consideration all the various factors that can affect an image quality. Thus, it is usually the case that an imperfect system is constructed to work well with some limited conditions, and only to revise it when more information is acquired. Therefore, in many practical situations, the presentation of training samples becomes sequential and adaptation to the system and forecasting by the system have to be alternatively carried out. In other practical applications (e.g. stock price forecasting), a static forecasting system is rendered ineffective by the changing property of the data source. Thus, a system is required to be incrementally adaptable to new information.

Progressive adaptation refers to the incremental modification of the network structure and parameters. The objective of progressive adaptation is to bring the network to produce a model close to a desired process through a gradual and progressive change of the network computational capability. In this paper, the progressive adaptation of a NFS is concerned with online learning of a data stream. The training exemplars from the data stream are assumed to arrive one at a time, and that the online learning algorithm adapts the structure and network parameters of the model based on the current training exemplar in an incremental manner. Online learning algorithms encompass the following four distinguishing features from *sequential learning algorithms* [15]: 1) All the training observations are sequentially (one-by-one) presented to the learning system; 2) At any time, only one training observation is seen and learned; 3) A training observation is discarded as soon as the learning procedure for that particular observation is completed; 4) The learning system has no prior knowledge as to how many total training observations will be presented. In addition, the following additional features are specifically considered for online learning algorithms: 1) There is no knowledge of the lower and upper bound of the training

observations; 2) The system should account for time-varying data characteristics (i.e. knowledge about data should change).

B. Focus and Structure of Paper

This paper focuses on the structure and learning aspects of a NFS [2] called *evolving Fuzzy Ensemble* (eFE). The proposed eFE network is organised into a modular and hierarchical network of neurons. Starting with zero neuron, eFE adopts a single-pass progressive learning approach that is inspired from the neurobiological learning theories of synaptic plasticity. This paper is organised as follows. Section II presents the architecture of the eFE network. Section III provides an overview of the proposed progressive biologically inspired learning algorithm for eFE network. In Sect. IV, two experiments are conducted to demonstrate the modelling and tracking capabilities of the proposed eFE model. Section V summarises and concludes this paper.

II. NETWORK ARCHITECTURE

The proposed eFE is a feed-forward network comprising of five layers of neurons. Other than layer I and V, the neurons in the other layers of eFE network are organised into a modular and hierarchical structure with H levels. Figure 1 shows the hierarchical structure of eFE network with two input variable nodes, one output variable node, and H hierarchical levels of nodes in layer II, III, and IV. The symbols used in this paper denotes the following:

I	number of input features;
M	number of output features;
K_h	number of rule nodes at h th hierarchical level;
IV_i	input linguistic node for input feature x_i ($1 \leq i \leq D$);
OV_m	output linguistic node for output feature y_m ($1 \leq m \leq M$);
$J_{h,i}$	number of linguistic term node/membership functions (MFs) for input feature x_i at h th hierarchical level;
$L_{h,m}$	number of MFs for output feature y_m at h th hierarchical level;
$IL_{h,i,j}$	j th MF for input feature x_i at h th hierarchical level ($1 \leq j \leq J_{h,i}$);
$OL_{h,m,\ell}$	ℓ th MF for output feature y_m at h th hierarchical level ($1 \leq \ell \leq L_{h,m}$);
$R_{h,k}$	k th rule node at h th hierarchical level ($1 \leq k \leq K_h$);
$C_{h,m,\ell,k}$	fuzzy rule connection between $R_{h,k}$ and $OL_{h,m,\ell}$.

The hierarchical organisation of eFE network corresponds to the structure of the neocortex [16]. A hierarchical level of eFE corresponds to a cortical layer. The differences in the sensory function among the cortical layers are simulated using different hierarchical levels of linguistic term/label nodes with different receptive fields (supports of fuzzy MFs). A layer of nodes in eFE network corresponds to a vertical cortical column. Every layer of neurons in the proposed eFE network is a functional module of a *fuzzy inference system* (FIS). Each layer of the eFE network performs specific fuzzy operations to realize an extended-Mamdani fuzzy reasoning

model that handles *mutual inclusivity* [17]. A fuzzy rule at the h th hierarchical level of eFE network is encoded as follows:

IF x_1 is $IL_{h,1,j} \dots \wedge \dots x_i$ is $IL_{h,i,j} \dots \wedge \dots x_I$ is $IL_{h,I,j}$ THEN y_1 is $OL_{h,1}, \dots, y_m$ is $OL_{h,m}, \dots, y_M$ is $OL_{h,M}$

where $OL_{h,m}$ represents a fuzzy partition that is shaped by the output MFs ($OL_{h,m,1}, \dots, OL_{h,m,L_{h,m}}$) of different degrees [18].

The input and output signals at any time t are respectively represented as crisp vectors $\mathbf{x}(t) = [x_1, \dots, x_I]^T$ and $\mathbf{y}(t) = [y_1, \dots, y_M]^T$. Every node in the first layer of the network takes in a single element from the input vector $\mathbf{x}(t)$. Every output node computes a single output element of the output vector $\mathbf{y}(t)$ with respect to the input stimulus vector $\mathbf{x}(t)$. The forward reasoning with signal propagation through each layer of the network as shown in Fig. 1 is described as follows.

A. Layer I: Input Linguistic Nodes

A node IV_i simply receives signal from the input stimulus x_i , and propagates it to the next layer nodes for processing. The output o_i^I of IV_i is defined in (1).

$$o_i^I = x_i. \quad (1)$$

B. Layer II: Input Linguistic Term Nodes

A neuron $IL_{h,i,j}$ denotes the j th fuzzy MF of the i th input feature at the h th hierarchical level and constitutes an antecedent of a fuzzy rule. Every node $IL_{h,i,j}$ defines a triangular fuzzy MF and its output $o_{h,i,j}^{II}$ is computed using *singleton fuzzification* operation in (2).

$$o_{h,i,j}^{II} = \mu_{h,i,j}^{II}(o_i^I; a_{h,i,j}^{II}, b_{h,i,j}^{II}, c_{h,i,j}^{II}) \\ = \max \left(\min \left(\frac{o_i^I - a_{h,i,j}^{II}}{b_{h,i,j}^{II} - a_{h,i,j}^{II}}, \frac{c_{h,i,j}^{II} - o_i^I}{c_{h,i,j}^{II} - b_{h,i,j}^{II}} \right), 0 \right), \quad (2)$$

where parameters $a_{h,i,j}^{II}$ and $c_{h,i,j}^{II}$ are the minimum and maximum support, and $b_{h,i,j}^{II}$ is the kernel of the MF for $IL_{h,i,j}$ respectively. In addition, the following conditions are satisfied for the input fuzzy MFs.

$$o_{h,i,1}^{II} = \begin{cases} 1, & o_i^I \leq b_{h,i,1}^{II}; \\ \mu_{h,i,1}^{II} \text{ defined in (2)}, & \text{otherwise.} \end{cases} \quad (3)$$

$$o_{h,i,J_{h,i}}^{II} = \begin{cases} 1, & o_i^I \geq b_{h,i,J_{h,i}}^{II}; \\ \mu_{h,i,J_{h,i}}^{II} \text{ defined in (2)}, & \text{otherwise.} \end{cases} \quad (4)$$

$$\sum_{j=1}^{J_{h,i}} \mu_{h,i,j}^{II} = 1; \quad b_{h,i,j}^{II} < b_{h,i,j+1}^{II}. \quad (5)$$

Equation (3) and (4) respectively defines the first and the last MF in $x_i \in \mathbb{R}$, and (5) specifies the overlap level of 0.5 between any two successive MFs. The eFE network consists of H sets (each set belonging to a level of eFE) of input linguistic labels, with each of them characterized by a scale parameter $\sigma_{h,i}^{II}$ defined in (6),

$$\sigma_{h,i}^{II} = b_{h,i,j}^{II} - a_{h,i,j}^{II} = c_{h,i,j}^{II} - b_{h,i,j}^{II}, \quad (6)$$

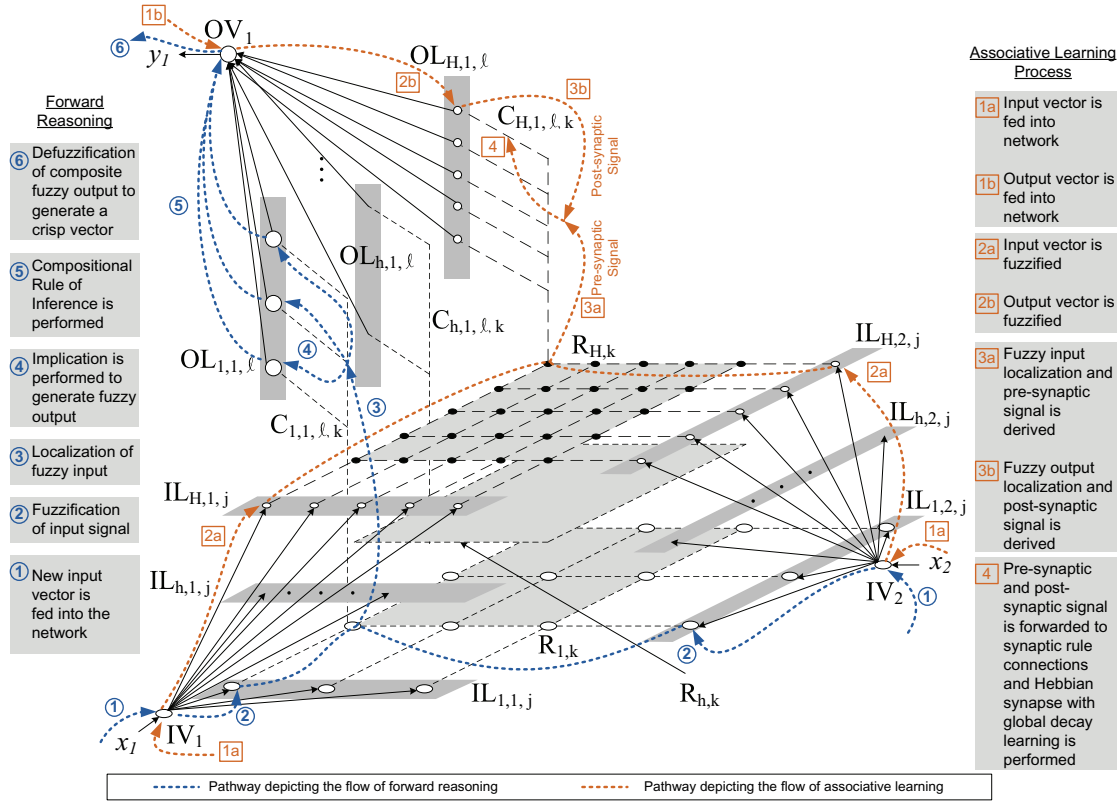


Fig. 1. Structure of eFE network.

and $\sigma_{h,i}^{II} > \sigma_{h+1,i}^{II}$. Each higher hierarchical level composes of MFs with smaller uniform entropies than the ones at the lower hierarchical level [19].

C. Layer III: Rule Nodes

A neuron $R_{h,k}$ represents the k th fuzzy rule node at the h th hierarchical level with output $o_{h,k}^{II}$ computed using a conjunctive fuzzy set operation (minimum t-norm operation) [20] defined in (7).

$$o_{h,k}^{II} = \min_{(h,i,j) \in A_{h,k}^{II}} o_{h,i,j}^{II}, \quad (7)$$

where $A_{h,k}^{II}$ is the set of input linguistic term nodes in layer II (antecedent) that are connected to $R_{h,k}$.

D. Layer IV: Output Linguistic Term Nodes

A neuron $OL_{h,m,\ell}$ represents the ℓ th output linguistic term node of the m th output linguistic node at the h th hierarchical level. Each node implements the *Compositional Rule of Inference* (CRI) [21] scheme. The output $o_{h,m,\ell}^{IV}$ of $OL_{h,m,\ell}$ is defined in (8).

$$o_{h,m,\ell}^{IV} = \max_{k \in R_{h,m,\ell}^{III}} (\min(o_{h,k}^{III}, C_{h,m,\ell,k}^f)), \quad (8)$$

where $R_{h,m,\ell}^{III}$ is the set of rule nodes in layer III that are connected to $OL_{h,m,\ell}$, and $C_{h,m,\ell,k}^f$ (refer to (13)) is the certainty factor (normalized weight) of the rule node connection $C_{h,m,\ell,k}$ between $R_{h,k}$ and $OL_{h,m,\ell}$.

The eFE network consists of H sets of output linguistic labels. These labels also satisfy the same conditions ((3), (4), (5)) that are specified for the input linguistic labels. Each output linguistic label is characterized by a scale parameter $\sigma_{h,m}^{IV}$ defined in (9),

$$\sigma_{h,m}^{IV} = b_{h,m,\ell}^{IV} - a_{h,m,\ell}^{IV} = c_{h,m,\ell}^{IV} - b_{h,m,\ell}^{IV}, \quad (9)$$

and $\sigma_{h,m}^{IV} > \sigma_{h+1,m}^{IV}$. At each hierarchical level, the overlap of 0.5 between any two successive output MFs implies no reconstruction error in the linguistic-numerical transformation [19].

E. Layer V: Output Linguistic Nodes

The neuron or defuzzification node OV_m defuzzifies the fuzzy composite output generated in layer IV into a crisp value. The output o_m^V of OV_m is computed using the *height defuzzification* method [20] defined in (10).

$$o_m^V = y_m = \frac{\sum_{h=1}^H \sum_{\ell=1}^{L_{h,m}} o_{h,m,\ell}^{IV} b_{h,m,\ell}^{IV}}{\sum_{h=1}^H \sum_{\ell=1}^{L_{h,m}} o_{h,m,\ell}^{IV}}, \quad (10)$$

where $b_{h,m,\ell}^{IV}$ represents the kernel of the MF that is employed to describe $OL_{h,m,\ell}$.

III. PROGRESSIVE LEARNING ALGORITHM

Figure 2 depicts the overview of the progressive learning procedure. The learning algorithm consists of three modules; namely: *network expansion*, *associative learning*, and *network*

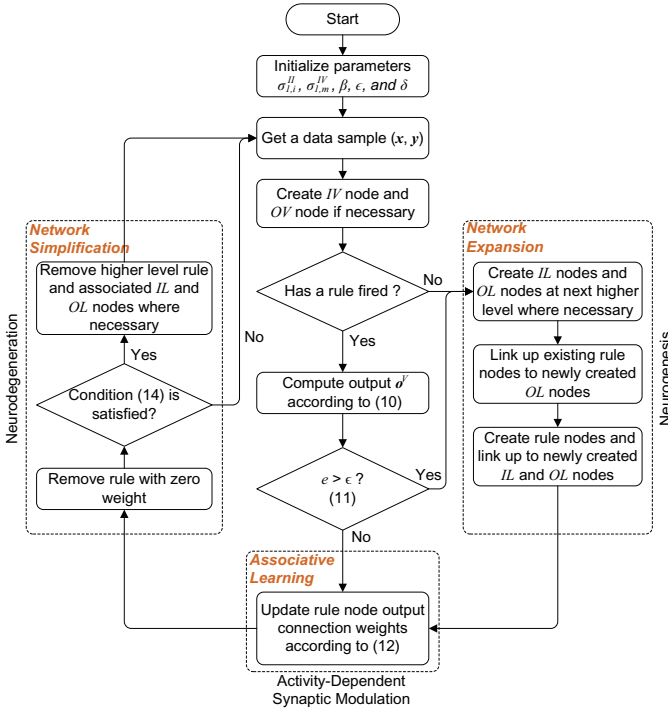


Fig. 2. Progressive learning procedure.

simplification. The learning procedure for each of the module is outlined as follows.

A. Network Expansion Module

The biological network structure plasticity is partially due to *neurogenesis* [22], [23], [24]. The function of neurogenesis was hypothesized to avoid prediction error due to *catastrophic interference* [24] by extending the encoding of new pattern with new neuron. In addition, microscopic examination of a coronal section through the neocortex shows that the neuronal cell bodies are arranged in different layers [25]. Birth dating studies showed that cortical neurons generated earlier reside in deeper layers while later-born neurons migrates past existing layers to form superficial layers [25].

First, to account and model the neurogenesis process and development in neocortex, new higher level rule node and its associated linguistic label nodes (IL and OL nodes) are added to the proposed eFE network whenever the network has a forecast error $e(t)$ exceeding a maximum acceptable error threshold ϵ defined in (11).

$$e(t) = \|\mathbf{o}^V(t) - \mathbf{y}^d(t)\|/M > \epsilon, \quad (11)$$

where $\mathbf{o}^V(t)$ is the output vector produced by eFE, and $\mathbf{y}^d(t)$ represents the desirable response the network should produce at time t .

The incremental insertion of new level neurons corresponds to the sequence of neurogenesis process in neocortex, in which neurons that are generated earlier reside in the deeper cortical layers (lower levels of eFE network). In addition, new neurons are added to avoid catastrophic interference and reduce prediction error according to the function of neurogenesis [24].

Second, new input and output linguistic label nodes are added with smaller fuzzy membership “width” (smaller support) at each higher level. Every new higher level linguistic label node is created with receptive field half the size of the immediate lower level linguistic label node. This creates linguistic label nodes in a hierarchical organisation with different receptive fields to simulate the laminar differences in the sensory function among the different cortical layers [16].

As a result, linguistic label nodes with smaller scales (receptive fields) will be present in those regions where the most subtle details are located. Nodes are adaptively allocated with adequate scale in the different regions of the input and output data space to form a sparse approximation of the data. The differences in the sensory function encoded by the different level of linguistic label nodes also allow the proposed network to describe the input and output feature across different scales. The lower level neurons could encode knowledge on more general information, and the higher level neurons to describe knowledge in finer detail.

Third, neurons created earlier (having wider receptive fields in lower levels) will be activated more frequently than higher level neurons. Consistent with physiological findings, in which older neurons are relatively more stable than newer neurons [26], this mechanism allows the lower and older neurons less likelihood from being removed (cf. Sect. III-C). In addition, this mechanism is also consistent with memory retention behaviour in human. Human normally remember more general than specific fact/knowledge.

Fourth, individually adding node as required limit the network complexity by preventing network to memorize training patterns and generalize poorly (large variance case). The network is also adapted according to changing pattern statistics to prevent the network being too simple that it can not even be able to represent training patterns well (large bias case) [27].

B. Associative Learning Module

Numerous evidences have suggested that rehearsal is required to support the consolidation of memories [28], [29], [30]. At the molecular level, the formation and consolidation of *long-term memory* are thought to be ultimately expressed in the form of structural changes at the synapses [31]. At the same time, other studies suggested that the decay of memory traces is the result of an active forgetting mechanism which opposes the strengthening mechanism involved in memory persistence [32], [33]. The progressive learning algorithm (cf. the associative learning process depicted in Fig. 1) captures these two key features; namely: retention and forgetting of memory traces, as follows.

- *Retention Mechanism*: First, to model the involvement of retention mechanism in long-term memory formation, the dynamics of the Hebbian learning rule [34] is incorporated to mediate the synapses between rule nodes and output linguistic term nodes. Synapse is strengthened by the retention term in (12) during correlated activation (firing) between rule node (pre-synaptic neuron) and output linguistic term node (post-synaptic neuron).

- *Forgetting Mechanism*: Second, the forgetting mechanism is interpreted as having a destabilizing effect on the existing stored memories. To account for such a process, a global synaptic decay term is included in (12) to weaken the synaptic strength.

The overall changes in the synaptic strength (Step 4 of the associative learning process depicted in Fig. 1) is simulated in (12),

$$\Delta C_{h,m,\ell,k}^w(t) = C_{h,m,\ell,k}^w(t) - C_{h,m,\ell,k}^w(t-1) \quad (12)$$

$$= \underbrace{\min(o_{h,k}^{III}(t), \mu_{h,m,\ell}^{IV}(y_m^d(t)))}_{\text{retention term}} - \underbrace{\beta \cdot C_{h,m,\ell,k}^w(t-1)}_{\text{forgetting term}},$$

where $C_{h,m,\ell,k}^w$ denotes the synapse strength of rule node connection $C_{h,m,\ell,k}$, $\mu_{h,m,\ell}^{IV}(y_m^d(t))$ represents the membership grade of $OL_{h,m,\ell}$ for output variable node OV_m (Step 2b of associative learning process in Fig. 1), $y_m^d(t)$ denotes the m th element of the desired output vector $\mathbf{y}^d(t)$ at time instance t , and β represents the forgetting rate.

To prevent saturation and loss of input selectivity, the weights of the connections between rule nodes and output linguistic term nodes are normalized according to (13).

$$C_{h,m,\ell,k}^f(t) = \frac{C_{h,m,\ell,k}^w(t)}{\sum_{\ell=1}^{L_{h,m}} C_{h,m,\ell,k}^w(t)}, \quad (13)$$

where $C_{h,m,\ell,k}^f$ is the normalized weight of rule node connection $C_{h,m,\ell,k}$. $C_{h,m,\ell,k}^f$ is the degree (certainty factor) of the output MF $OL_{h,m,\ell}$ associated with $R_{h,k}$. Different certainty factors to the output MFs are used to shape an alternative output fuzzy answer space [18].

The computational model demonstrates the effect of active forgetting and retention memory processes. Only in the presence of sufficient and repeated synaptic reinforcements, either triggered by multiple repetitive training events or reactivation/recollection of such events, can memories be maintained and strengthened. This is consistent with evidence, in which the more frequently an event is recalled, the better and longer it will be remembered [28].

C. Network Simplification Module

Even though higher level rules and associated linguistic label nodes are adaptively added, the consolidated rules will be updated according to changing pattern statistics. Some rule nodes that represent the different localized regions of the data space may at a later stage of adaptation make to produce similar forecast as those rules activated at a lower level.

At each presentation of a data sample, the highest level activated rule $R_{H_f,k}^f$ is removed if the following condition (14) is satisfied.

$$\|\mathbf{o}_{H_f,k}^R - \mathbf{o}_{H_f-1,a}^R\|/M < \delta; \exists a \in R_{H_f-1}^f, H_f \neq 1, \quad (14)$$

where δ represents the output forecast similarity threshold, H_f represents the highest level of the rules that are activated, $R_{H_f-1}^f$ represents the set of lower immediate level rules that are activated, and $\mathbf{o}_{H_f,k}^R$ represents the inferred crisp output

vector of $R_{H_f,k}^f$ using CRI scheme defined in (15) with $h = H_f$.

$$\mathbf{o}_{h,k}^R = [o_{h,k,1}^R, \dots, o_{h,k,m}^R, \dots, o_{h,k,M}^R]^T;$$

$$o_{h,k,m}^R = \frac{\sum_{\ell=1}^{L_{h,m}} \min(o_{h,k}^{III}, C_{h,m,\ell,k}^f) \cdot b_{h,m,\ell}^{IV}}{\sum_{\ell=1}^{L_{h,m}} \min(o_{h,k}^{III}, C_{h,m,\ell,k}^f)}. \quad (15)$$

Such higher level rule with smaller receptive field can safely be removed as it describes the localized fuzzy region that is also enclosed by a lower immediate level rule.

In addition, if a rule node $R_{h,k}$ has all its rule node connection weights $C_{h,m,\ell,k}^w$ ($1 \leq \ell \leq L_{h,m}$; $1 \leq m \leq M$) reduced to zero then $R_{h,k}$ is removed from the rule base. This is because the rule node is not frequently activated to make it relevant for the present data events. Thus, the network structure is simplified in this process. This step corresponds to structure plasticity related to the removal of neurons and synapses that contributes to the loss of memory (see, e.g., [35], for instance).

IV. EXPERIMENTS: SIMULATIONS

In this section, two experiments are conducted to demonstrate the quality features of the proposed eFE model.

A. Online System Identification of a Nonlinear System

The objective of this experiment is to evaluate the proposed eFE network in one-pass identification of a nonlinear system defined by (16).

$$y_d(t+1) = \frac{y_d(t)}{1 + y_d^2(t)} + u^3(t), \quad (16)$$

where $u(t) \in [-1.0, 1.0]$ is given by $\sin(2\pi t/100)$, and $y_d(t) \in [-1.5, 1.5]$ and $y_d(0) = 0$. The proposed eFE model is benchmarked with existing works; namely, MRAN [36], RANEKF [37], SONFIN [38], SAFIS [14], Simpl_eTS [39], eTS [13], DENFIS [40], FALCON-ART [41], and EFuNN [42] under the same experimental set-up in [14].

Table I lists the consolidated generalization result achieved by the various modelling techniques. The result shows that neural network and TSK modelling techniques deliver better generalization accuracies than the proposed eFE model. The following are the reasons to explain this result.

- TSK-based fuzzy models [43], [44] are used for *precise fuzzy modelling* (PFM) in place of *linguistic fuzzy modelling* (LFM) [45]. The proposed eFE model is based on Mamdani fuzzy model used for LFM.
- The benchmark modelling techniques (MRAN, RANEKF, SONFIN, SAFIS, Simpl_eTS, eTS, DENFIS, FALCON-ART, and EFuNN¹) utilize a clustering-based prototyping to partition the data space. In contrast, eFE makes use of the more restrictive grid based type of data space partitioning [46]. Clustering-based prototyping poses less

¹The second layer nodes in EFuNN are fully connected to a rule node in the third layer via a set of connection weights. These connection weights represent the coordinates of the center of the hypersphere in the fuzzy input space for a rule node.

constraint as this methodology provides direct modelling of data distribution in the data space. Hence, clustering-based partitioning of data space usually requires lesser number of prototypes (rule nodes). However, systems based on clustering-based prototyping usually have opaque computing structure. For example, the input fuzzy sets derived by SAFIS have large overlaps that obstruct the understandings of the underlying data structure (refer to [14]).

- The other benchmark systems employ complex operations and perform fine tuning of network parameters. The proposed eFE model does not require such time consuming complex operations. For example, RANEKF and MRAN networks are evolving radial basis neural networks based on the principle of the Resource Allocation Network (RAN) [47]. RANEKF extends the RAN by employing the Extended Kalman Filter (EKF) algorithm in place of the Least Mean Squares (LMS) algorithm. MRAN extends RAN (and RANEKF) by combining the growth criterion of the RAN with a pruning strategy based on the relative contribution of each hidden unit to the overall network output. SAFIS computes the statistical contribution of every rule to the system output and the parameters are tuned with EKF algorithm. eTS has to recursively update its TSK model structure based on the potential of the input data. This is defined based on its distances to all other data points that have been received. As for the DENFIS online model, a weighted recursive Least-Square Estimator (LSE) is used. In SONFIN, the consequent parameters of its TSK rules are tuned either by LMS or Recursive Least Squares (RLS) algorithms and the precondition parameters are tuned by back-propagation algorithm. FALCON-ART tunes the parameters of the MFs using the back-propagation algorithm. Unconstraint accuracy-oriented parameter optimization is also one of the other reasons that will produce significant overlapping of the fuzzy MFs. This obstructs the understandings of the underlying data structure.
- The training and testing data are generated from the same difference equation (16). An accuracy-oriented modelling technique with enough training cycles to tune its network structure and parameters would definitely perform better in this case. On the other hand, the eFE learning algorithm does not try to reduce the error to zero. Under soft computing paradigm, the proposed eFE aims to contain the error between the computed output and desirable values within an acceptable threshold. Hence, the learning error is not asymptotically reduced to zero. It is seldom the case in reality that the unseen events replicate exactly the same as past events. Hence, optimization done at the present time instance has no guarantee in providing an optimal solution that will also generalize well for future unseen events. For one obvious reason, current optimization may produce a solution that over fits to the current set of training data. Section IV-B illustrates this point by evaluating EFuNN, DENFIS and eFE to track a

TABLE I
FORECAST RESULTS ON ONLINE NONLINEAR SYSTEM IDENTIFICATION

Reasoning Model	Type	# Rules	Testing RMSE
MRAN ^a	NN	10*	0.0129
RANEKF ^a	NN	11*	0.0184
SONFIN ^a	TSK	10	0.0130
SAFIS ^a	TSK	8	0.0116
Simpl_eTS ^a ($r=0.075, \Omega=10^6$)	TSK	18	0.0122
eTS ^a ($r=1.0, \Omega=10^6$)	TSK	19	0.0082
DENFIS	TSK	8	0.0482
FALCON-ART	M	289	0.1379
EFuNN	M	18	0.0578
eFE	E-M	21	0.0751

^a = Result adapted from [14]; NN = Neural Network; TSK = Takagi-Sugeno-Kang-based fuzzy model; M = Mamdani-based fuzzy model; E-M = Extended Mamdani-based fuzzy model that handles mutual inclusivity [17]; * = Number of hidden nodes; RMSE = Root Mean Squared Error.

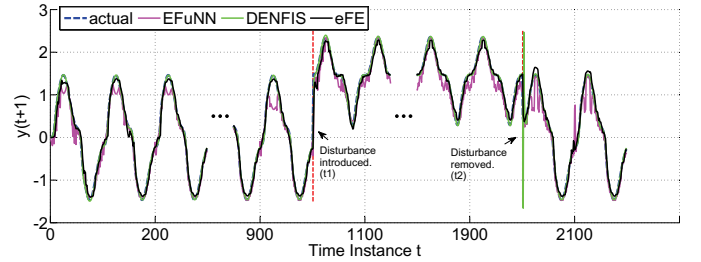


Fig. 3. Forecast of time-varying nonlinear system.

time-varying nonlinear system.

B. Online System Identification of a Time-Varying Nonlinear System

The previous nonlinear dynamic system (16) is modified as shown in (17).

$$y_d(t+1) = \frac{y_d(t)}{1 + y_d^2(t)} + u^3(t) + f(t), \quad (17)$$

where $f(t)$ simulates the disturbance to be introduced, and is defined in (18),

$$f(t) = \begin{cases} 0, & 0 \leq t \leq 1000 \text{ and } t \geq 2001; \\ 1, & 1001 \leq t \leq 2000. \end{cases} \quad (18)$$

Similar to [48], the eFE model is employed to progressively identify the characteristics of the modified nonlinear dynamic system.

Figure 3 shows the tracking capability of eFE, EFuNN, and DENFIS. The eFE model is able to track the underlying system characteristics right from the initial stage. The proposed eFE model is also able to track new system characteristics when the disturbance $f(t)$ is introduced at time t_1 ($t = 1001$) and when it is removed at time t_2 ($t = 2001$). EFuNN has large tracking error towards the end of the simulation. DENFIS exhibits large tracking error at time t_2 .

Figure 4(a) illustrates the corresponding changes to the structure of the three models. For eFE, rule nodes are progres-

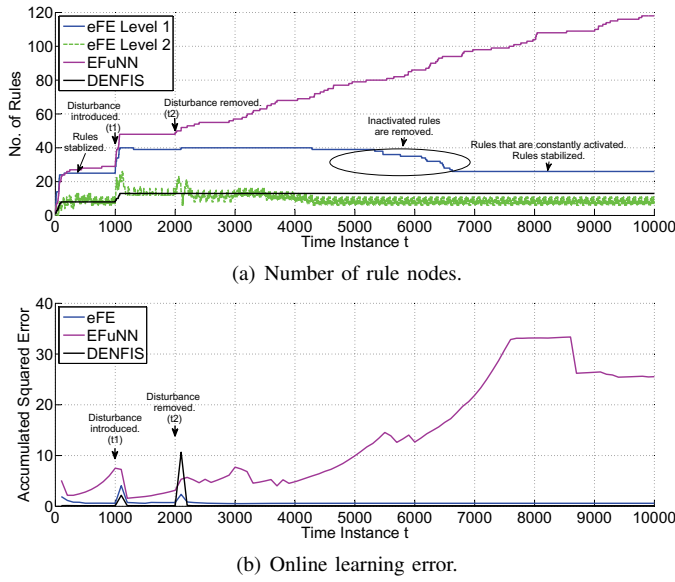


Fig. 4. Online learning of the time-varying nonlinear system.

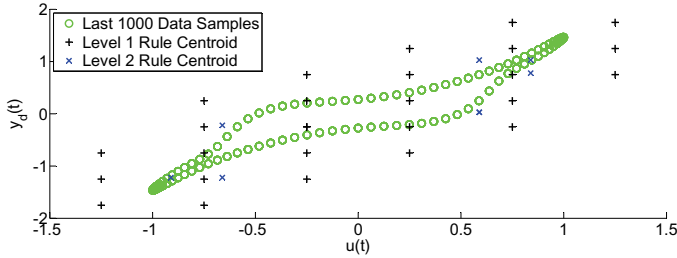


Fig. 5. Rules identified by eFE at the end of simulation ($t = 10000$).

sively added and removed. After the disturbance is removed at t_2 , the fuzzy rules that describe the disturbance are subsequently removed. This is because these rules are not activated frequently enough to make it relevant for the present data events. For EFuNN, rule nodes are progressively added even when disturbance has been removed at t_2 . The structure of DENFIS is fixed after the introduction of disturbance at time t_1 .

Figure 4(b) illustrates the accumulated squared error computed over 100 adjacent time steps. Spikes in the eFE learning error are observed at three instances; namely, during the initial stage at time $t = 0$, when the disturbance is introduced at time t_1 , and when the disturbance is removed at t_2 . The increase in learning error at these three instances corresponds to the time when there are changes in the underlying properties of the nonlinear system model. When the disturbance $f(t)$ is removed at t_2 only a smaller increase in eFE learning error has been observed. In contrast, EFuNN performs unacceptably. The online learning error of EFuNN generally becomes larger at each time instance. On the other hand, DENFIS has overly learnt and fitted the data. This is because the online learning error accumulated by DENFIS just after the disturbance was removed, is higher than the amount of error accumulated when the disturbance was introduced. The information learnt by

DENFIS before the disturbance was introduced is lost even though the number of rules has not reduced. This experiment illustrates that DENFIS accuracy-oriented optimization on the given information causes the network to quickly skew towards current information and forget the information learnt in the past.

Figure 5 plots the distribution of the last 1000 input data samples, and the centroids of the rules identified by eFE. The figure shows more rules are identified near the lower left and upper right region of the input data space. These regions correspond to the input data samples that are more closely located. The lower level eFE rules coarsely describe these regions and higher level rules are added to describe these regions in finer detail.

V. CONCLUSION

This paper presents an online learning-based hierarchical NFS called, eFE (evolving Fuzzy Ensemble). A progressive learning algorithm is proposed to incrementally modify the hierarchical structure of eFE in accordance with neurophysiological studies. With continuous exposure to external stimuli, the eFE network adapts its structure by allocating nodes and creating links to autonomously support fuzzy data associations. Activity-dependent synapse with global decay learning rule is incorporated such that fuzzy rules that are relevant to the observed training data are reinforced while rules that are inconsistent are weakened. This technique simulates the two opposite mechanisms that are involved in memory persistence [32], [33], [28]. At the same time, the eFE network structure is simplified by removing similar and irrelevant fuzzy knowledge. This is accomplished by removing rules that have similar forecast output, and irrelevant rules that are no longer reinforced by the external stimuli. This technique corresponds to structure plasticity related to the removal of neurons and synapses that contributes to the loss of memory (see, e.g., [35], for instance).

The learning algorithm employed by eFE possesses the distinguishing features of online learning algorithms as outlined in Sect. I. The experimental results are concluded as follows.

- The experimental result showed that the proposed eFE model is able to adequately model the underlying data characteristics of a nonlinear system compared to other benchmarked modelling techniques.
- The tracking capability of eFE model was demonstrated by learning the underlying data characteristics of a time-varying system. Without prior knowledge of the data distribution, the eFE multi-resolution hierarchy structure allows both good generalization and fine detail to be captured. Unlike other benchmark modelling techniques, the experiment also showed the capability of the proposed eFE in removing irrelevant fuzzy knowledge. The experiment also demonstrated that optimization can cause a model to quickly skew towards current information and forget the information learnt in the past.

In conclusion, most batch learning methodologies are computationally intensive with the objective of deriving a global

solution by optimizing certain criterion (e.g. parameter adaptation to minimize prediction error under supervised learning). The extension of offline learning paradigms to online learning mode questions the need for such intensive computation given that only limited information is available at a given time. After all, such form of *limited-knowledge optimization* in time-varying environment would likely not produce a globally optimized solution which directly contradicts the initial idea for optimization in a global sense. Instead, this paper presents a simple fuzzy associative learning algorithm that does not rely on computationally intensive accuracy-oriented optimization to incrementally adapt an ensemble of fuzzy models.

REFERENCES

- [1] L. A. Zadeh, "Fuzzy logic, Neural networks, and Soft Computing," *Communications of the ACM*, vol. 37, no. 3, pp. 77–84, 1994.
- [2] C. T. Lin and C. S. G. Lee, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*. Englewood Cliffs, NJ: Prentice Hall, 1996.
- [3] M. E. Yüksel and E. Beşdok, "A Simple Neuro-Fuzzy Impulse Detector for Efficient Blur Reduction of Impulse Noise Removal Operators for Digital Images," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 6, pp. 854–865, 2004.
- [4] C. W. Ting and C. Quek, "A Novel Blood Glucose Regulation Using TSK⁰-FCMAC: A Fuzzy CMAC based on the Zero-Ordered TSK Fuzzy Inference Scheme," *IEEE Transactions on Neural Networks*, vol. 20, no. 5, pp. 856–871, 2009.
- [5] K. K. Ang and C. Quek, "RSPOP: Rough Set-Based Pseudo Outer-Product Fuzzy Rule Identification Algorithm," *Neural Computation*, vol. 17, no. 1, pp. 205–243, 2005.
- [6] G. Castellano, A. Fanelli, and C. Mencar, "A Neuro-Fuzzy Network to generate Human-Understandable Knowledge from Data," *Cognitive Systems Research*, vol. 3, no. 2, pp. 125–144, 2002.
- [7] E. Y. Cheu, S. K. Ng, and C. Quek, "Enhancing Decision Support System with Neural Fuzzy Model and Simple Model Visualizations," in *Handbook on Decision Making, ISRL 4*, L. C. Jain and C. P. Lim, Eds. Springer-Verlag Berlin Heidelberg, 2009, ch. 4, pp. 85–116.
- [8] W. Wang, F. Ismail, and F. Golnaraghi, "A Neuro-Fuzzy Approach to Gear System Monitoring," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 5, pp. 710–723, 2004.
- [9] K. Kiguchi, T. Tanaka, and T. Fukuda, "Neuro-Fuzzy Control of a Robotic Exoskeleton with EMG Signals," *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 4, pp. 481–490, 2004.
- [10] J. Kim and N. Kasabov, "HyFIS: Adaptive Neuro-Fuzzy Inference Systems and their Application to Nonlinear Dynamical Systems," *Neural Networks*, vol. 12, no. 9, pp. 1301–1319, 1999.
- [11] J. Sim, W. L. Tung, and C. Quek, "FCMAC-Yager: A Novel Yager-Inference-Scheme-Based Fuzzy CMAC," *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1394–1410, Nov. 2006.
- [12] W. Yu, F. O. Rodriguez, and M. A. Moreno-Armendariz, "Hierarchical Fuzzy CMAC for Nonlinear Systems Modeling," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1302–1314, 2008.
- [13] P. Angelov and D. P. Filev, "An Approach to Online Identification of Takagi-Sugeno Fuzzy Models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 1, pp. 484–498, 2004.
- [14] H.-J. Rong, N. Sundararajan, G.-B. Huang, and P. Saratchandran, "Sequential Adaptive Fuzzy Inference System (SAFIS) for Nonlinear System Identification and Prediction," *Fuzzy Sets and Systems*, vol. 157, no. 9, pp. 1260–1275, May 2006.
- [15] G. B. Huang, P. Saratchandran, and N. Sundararajan, "A Generalized Growing and Pruning RBF (GGAP-RBF) Neural Network for Function Approximation," *IEEE Transactions on Neural Networks*, vol. 16, no. 1, pp. 57–67, Jan. 2005.
- [16] J. A. Hirsch and L. M. Martinez, "Laminar Processing in the Visual Cortical Column," *Current Opinion in Neurobiology*, vol. 16, no. 4, pp. 377–384, Aug. 2006.
- [17] J. Tan and C. Quek, "ACPOP: Ambiguity Correction-based Pseudo-Outer-Product Fuzzy Rule Identification Algorithm," in *IEEE International Conference on Fuzzy Systems*, Aug. 2009, pp. 74–79.
- [18] W. Siler and J. J. Buckley, *Fuzzy Expert Systems and Fuzzy Reasoning*. Hoboken, NJ: Wiley, 2005.
- [19] W. Pedrycz, "Why Triangular Membership Functions?" *Fuzzy Sets and Systems*, vol. 64, no. 1, pp. 21–30, 1994.
- [20] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ: Prentice Hall, 2001.
- [21] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [22] R. A. Chambers, M. N. Potenza, R. E. Hoffman, and W. Miranker, "Simulated Apoptosis/Neurogenesis Regulates Learning and Memory Capabilities of Adaptive Neural Networks," *Neuropsychopharmacology*, vol. 29, pp. 747–758, 2004, anglais.
- [23] S. Becker, "A Computational Principle for Hippocampal Learning and Neurogenesis," *Hippocampus*, vol. 15, no. 6, pp. 722–738, 2005.
- [24] L. Wiskott, M. J. Rasch, and G. Kempermann, "A Functional Hypothesis for Adult Hippocampal Neurogenesis: Avoidance of Catastrophic Interference in the Dentate Gyrus," *Hippocampus*, vol. 16, no. 3, pp. 329–343, 2006, automatic medline import.
- [25] J. G. Parnavelas, "The Origin of Cortical Neurons," *Brazilian Journal of Medical and Biological Research*, vol. 35, pp. 1423–1429, 2002.
- [26] C. Schmidt-Hieber, P. Jonas, and J. Bischofberger, "Enhanced Synaptic Plasticity in Newly Generated Granule Cells of the Adult Hippocampus," *Nature*, vol. 429, no. 6988, pp. 184–187, May 2004.
- [27] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley Interscience, 2000.
- [28] J. Jonides, R. L. Lewis, D. E. Nee, C. A. Lustig, M. G. Berman, and K. S. Moore, "The Mind and Brain of Short-Term Memory," *Annual Review of Psychology*, vol. 59, pp. 193–224, 2008.
- [29] J. L. C. Lee, "Reconsolidation: Maintaining Memory Relevance," *Trends in Neurosciences*, vol. 32, no. 8, pp. 413–420, 2009.
- [30] B. A. Kuhl, A. T. Shah, S. DuBrow, and A. D. Wagner, "Resistance to Forgetting Associated with Hippocampus-Mediated Reactivation During New Learning," *Nature Neuroscience*, vol. 13, no. 4, pp. 501–506, Apr. 2010.
- [31] E. R. Kandel, "The Molecular Biology of Memory Storage: A Dialogue Between Genes and Synapses," *Science*, vol. 294, no. 5544, pp. 1030–1038, 2001.
- [32] M. C. Anderson, K. N. Ochsner, B. Kuhl, J. Cooper, E. Robertson, S. W. Gabrieli, G. H. Glover, and J. D. E. Gabrieli, "Neural Systems Underlying the Suppression of Unwanted Memories," *Science*, vol. 303, no. 5655, pp. 232–235, Jan. 2004.
- [33] M. Johansson, A. Aslan, K.-H. Bauml, A. Gabel, and A. Mecklinger, "When Remembering Causes Forgetting: Electrophysiological Correlates of Retrieval-Induced Forgetting," *Cerebral Cortex*, vol. 17, no. 6, pp. 1335–1341, Jun. 2007.
- [34] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. New York: John Wiley & Sons Inc., 1949.
- [35] J. H. Morrison and P. R. Hof, "Life and Death of Neurons in the Aging Brain," *Science*, vol. 278, no. 5337, pp. 412–419, Oct. 1997.
- [36] Y. Lu, N. Sundararajan, and P. Saratchandran, "A Sequential Learning Scheme for Function Approximation using Minimal Radial Basis Function Neural Networks," *Neural Computation*, vol. 9, no. 2, pp. 461–478, 1997.
- [37] V. Kadirkamanathan and M. Niranjan, "A Function Estimation Approach to Sequential Learning with Neural Networks," *Neural Computation*, vol. 5, no. 6, pp. 954–975, 1993.
- [38] C. F. Juang and C. T. Lin, "An On-line Self-Constructing Neural Fuzzy Inference Network and its Applications," *IEEE Transactions on Fuzzy Systems*, vol. 6, no. 1, pp. 12–32, 1998.
- [39] P. Angelov and D. P. Filev, "SimplEFS: A Simplified Method for Learning Evolving Takagi-Sugeno Fuzzy Models," in *IEEE International Conference on Fuzzy Systems*, Ford Motor Co., Glendale Ave 24500, Detroit, United States, 2005, pp. 1068–1073.
- [40] N. Kasabov and Q. Song, "DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and its Application for Time-Series Prediction," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 144–154, April 2002.
- [41] C. J. Lin and C. T. Lin, "An ART-based Fuzzy Adaptive Learning Control Network," *IEEE Transactions on Fuzzy Systems*, vol. 5, no. 4, pp. 477–496, 1997.
- [42] N. Kasabov, "Evolving Fuzzy Neural Networks for Supervised/Unsupervised Online Knowledge-Based Learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 31, no. 6, pp. 902–918, 2001.
- [43] T. Takagi and M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 1, pp. 116–132, February 1985.
- [44] M. Sugeno and G. T. Kang, "Fuzzy Modelling and Control of Multilayer Incinerator," *Fuzzy Sets and Systems*, vol. 18, no. 3, pp. 329–345, 1986.
- [45] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, "Interpretability Improvements to Find the Balance Interpretability-Accuracy in Fuzzy Modeling: An Overview," in *Interpretability Issues in Fuzzy Modeling, Studies in Fuzziness and Soft Computing*, J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, Eds. Springer-Verlag New York, Inc., 2003, pp. 3–22.
- [46] C. T. Sun, "Rule-base Structure Identification in An Adaptive-Network-Based Fuzzy Inference System," *IEEE Transactions on Fuzzy Systems*, vol. 2, no. 1, pp. 64–73, 1994.
- [47] J. Platt, "A Resource Allocating Network for Function Interpolation," *Neural Computation*, vol. 3, no. 2, pp. 213–225, 1991.
- [48] C. F. Juang and Y. W. Tsao, "A Self-Evolving Interval Type-2 Fuzzy Neural Network With Online Structure and Parameter Learning," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1411–1424, 2008.