PANFIS: A Novel Incremental Learning Machine

Mahardhika Pratama, *Student Member, IEEE*, Sreenatha. G. Anavatti,
Plamen. P. Angelov, *Senior Member, IEEE*, and Edwin Lughofer

***Abstract*— Most of the dynamics in real-world systems are compiled by shifts and drifts, which are uneasy to be overcome by omnipresent neuro-fuzzy systems. Nonetheless, learning in nonstationary environment entails a system owning high degree of flexibility capable of assembling its rule base autonomously according to the degree of nonlinearity contained in the system. In practice, the rule growing and pruning are carried out merely benefiting from a small snapshot of the complete training data to truncate the computational load and memory demand to the low level. An exposure of a novel algorithm, namely parsimonious network based on fuzzy inference system (PANFIS), is to this end presented herein. PANFIS can commence its learning process from scratch with an empty rule base. The fuzzy rules can be stitched up and expelled by virtue of statistical contributions of the fuzzy rules and injected datum afterward. Identical fuzzy sets may be alluded and blended to be one fuzzy set as a pursuit of a transparent rule base escalating human's interpretability. The learning and modeling performances of the proposed PANFIS are numerically validated using several benchmark problems from real-world or synthetic datasets. The validation includes comparisons with state-of-the-art evolving neuro-fuzzy methods and showcases that our new method can compete and in some cases even outperform these approaches in terms of predictive fidelity and model complexity.**

***Index Terms*— Evolving neuro-fuzzy systems (ENFSs), incremental learning, sample-wise training.**

I. Introduction

A. Preliminary

THE salient motivation behind the use of fuzzy system is that it allows operations interpretable in a way akin to the human's logical reasoning. Fuzzy logic system proposed by Zadeh [1] is intelligible using fuzzy linguistic rule and can realize approximate reasoning to deal with imprecision and uncertainty in a decision-making process. These traits are preferable when the system is too complex to be analyzed by a physics-based approach or the source of information can be merely interpreted qualitatively, inexactly, or uncertainly. Traditional approaches in designing the fuzzy system are unfortunately overdependent on expert knowledge [2] and usually necessitate tedious manual interventions. This for brevity leads to a static rule base, which cannot be tuned once its initial setting to gain a better performance. The designers noticeably have to spend a laborious time to examine all input–output relationships of a complex system to elicit a representative rule base constraining its practicability in evolving dynamic and time critical environments.

This issue has led to the development of neuro-fuzzy systems (NFSs) [42], a powerful hybrid modeling approach that assimilates the learning ability, parallelism, and robustness of neural networks with the human-like linguistic and approximate reasoning traits of the fuzzy logic systems. The complex and dynamic natures of real-world engineering problems are in general complicated by time-varying or regime shifting issues. Classical NFSs are in contrast trained completely from offline data and remain as static models, which are impractical for nonstationary environments.

Recently, coping with nonstationary environments has drawn intensive research works for typical NFSs, which are able to adapt their parameters and to automatically expand their design contexts simultaneously. Evolving NFS (ENFS) based on the concept of incremental learning [3] accordingly opens a new unchartered territory as a plausible solution provider to settle time-varying cases that convey regime shifting and drifting properties, and enriched a landscape of fuzzy system for coping with time-variant systems in real-time fashions. The principal construct of the ENFS initiates a favorable cornerstone in handling time-varying system. That is, whenever a new characteristic of the system appears, the existing rules will automatically reorganize its structure or even split a supplementary rule to accommodate the uncovered data distributions. The creation of an extraneous rule in the following signifies the presence of the untouched data region, which could be a new characteristic of the process or a reaction to a new disturbance. This obviously produces a promising impetus to handle the nonstationary, shifting, and drifting effects of data streams [9] in real-time with immense opportunities for future smart devices and algorithms embedded on them.

B. Survey Over State-of-the-Art Works

In the early development of the ENFSs, most of the models are generally featureless as a complete dataset at hand *a priori* is indispensable for them to properly accomplish a given task dubbed as a batch learning scheme. A retraining

Manuscript received April 13, 2012; revised May 1, 2013; accepted June 22, 2013. Date of publication July 25, 2013; date of current version December 13, 2013. This work was supported in part by the Austrian fund for promoting scientific research under Contract I328-N23, the IREFS, the research program with the Austrian Center of Competence in Mechatronics, and the COMET K2 program of the Austrian government.

M. Pratama and S. G. Anavatti are with the School of Engineering and Information Technology, University of New South Wales, Australian Defence Force Academy, Canberra ACT 2600, Australia (e-mail: pratama@ieee.org; s.anavatti@adfa.edu.au).

P. P. Angelov is with the School of Computing and Communications, Lancaster university, Lancaster LA14WA, U.K. (e-mail: p.p.angelov@lancaster.ac.uk).

E. Lughofer is with the Department of Knowledge-Based Mathematical Systems, Johannes Kepler University of Linz, Linz A-4040, Austria (e-mail: edwin.lughofer@jku.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2013.2271933

phase with the use of an up-to-date dataset ought to be reciprocally executed whenever a new knowledge needs to be incorporated. A progressive learning of the new pattern will in essence modify the initial trained model in such a way that the previously learned but possibly still valid knowledge is completely erased. That is, the newly appearing information catastrophically discarded the model's memory of the previous learned knowledge because of its fixed learning capacity [8]. Dynamic fuzzy neural network (DFNN), generalized DFNN, and self-organizing fuzzy neural network were lodged as the solution providers for the aforementioned drawbacks. These ENFSs do not necessitate a complete dataset to be available at hand; yet, they gather the preceding training stimuli and in turn reuse them in the next training cycles. Such strategy intrinsically retards model updates and ensues an excessive utilization of memory capacity, which is not in line with the crux of efficient learning machine especially in facing a system with rapid varying attributes and a vast amount of training data. The solely realistic likelihood to obviate this demerit is to merely execute a single sample in every training episode—as achieved, e.g., in [11] and [12]. Parameter adaptation and rule evolution are settled in a sample-wise or single-pass manner, i.e., a single sample is loaded, the fuzzy model is updated with respect to this data point and the loaded sample is immediately discarded, afterward, and so on. This warrants prompt updates with low virtual memory usage.

Apart from automatically proliferating fuzzy rules, an ENFS is much more desirable to be capable of evicting inactive components (neurons and rules), which are no longer descriptive to reflect newly acquired data trends so as to strike a coherent tradeoff between the predictive accuracy and the model simplicity. This deficiency is indeed suffered from evolving Takagi–Sugeno (eTS), simplified eTS (simp_eTS), flexible fuzzy inference system (FLEXFIS) [11], [12], and [19] disabling to switch off superfluous fuzzy rules. By extension, in the noisy environment, there arises an avenue of a training data point, which may be outlier, is wrongly recruited as a new rule. Fortunately, such cluster is usually occupied with few data points, thus contributing little during its lifespan. A mismatch in generating a new rule can be counterbalanced with the so-called rule base simplification strategy for the sake of a compact and parsimonious rule base. In addition, an overcomplex rule base is always conjectured as a subject of an overfitting issue deteriorating the generalizing ability of a model and conceals rule semantics (i.e., considering a rule base with few hundred rules). Another paramount prerequisite of an effective ENFS to remedy these bottlenecks is an *ad hoc* manner in eliminating inconsequential rules mitigating its structural complexity affirming its robustness to an oversized rule base or is not greedy in crafting complementary rules while retaining its best predictive accuracy [13], [14].

On the one side, a major technical flaw in most of the ENFSs is induced by unidimensional membership functions generating hyperspherical clusters. One can approbate that self-organizing fuzzy modified least square, sequential adaptive fuzzy inference system (SAFIS), and growing–pruning fuzzy neural network [13], [39], [40] engage this type of fuzzy system. The rule evolved by this type of membership function cannot reflect real data distributions properly and imposes upper and lower bounds of training data to be available before the learning process commences. On the other side, another major shortcoming is a fact that ubiquitous approaches in devising EFNS benefit a product or a t-norm operator. This approach is deemed deficient as it does not underpin the possible input variable interactions. To remedy this bottleneck, a normalized distance (a Mahalanobis distance) is one of the recipes as it emphasizes the possible input variable interactions [15]. We can furthermore envisage multidimensional membership functions stimulating a more appealing property delineating ellipsoidal clusters in any directions, whose axes are not necessarily parallel to the input variable axes [14] where it is compatible with real-time or online requirements.

Another pivotal shortcoming of the state-of-the-art ENFSs is prone with outliers, which is exemplified by dynamic evolving neuro-fuzzy inference system (DENFIS), self-constructing neuro-fuzzy inference network [34], [45]. This is engendered by these approaches that solely pay attention to the distance of the fuzzy rules to the newest datum making use of the firing strength of the fuzzy rule or Euclidean distance improbably undoing the likelihood that the outliers are unified to be the fuzzy rules as the outliers can be distant or outside to the zone of influence. The so-called Learn++ algorithm was proposed in [49], which is in substance adaboost-like algorithm. It was expanded in [50] to deal with dynamic number of target class and in [51] to cope with regime drifting nature of training samples. The state-of-the-art algorithm of this machine learning type was pioneered by Street and Kim [56] with streaming ensemble algotirhm. Another prominent work, namely dynamic weighted majority, was proposed by Kolter and Maloof [57]. It is as with Learn++ devising a passive drift detection. It is worth stressing that Learn++ family emphasizes the concept of ensemble and parsimonious network based on fuzzy inference system (PANFIS) is obviously capable of serving as base learner in Learn++ working framework.

The evolving algorithm was developed under the framework of type-2 fuzzy system in [52]. Reference [53] outlines the extended version of [52] not only in type-2 fuzzy system environment, but also in recurrent network topology. Research work [54] exhibits the implementation of EFS in the embedded system. The variant of EFS to the classification cases was presented in [55]. Note that the enhanced versions of PANFIS in type-2 fuzzy system, recurrent network topology, or classification problems are the subjects of future investigations.

C. Proposed Algorithm

A seminal ENFS namely PANFIS is proposed herein. PANFIS is likewise capable of starting its rehearsal process from scratch with an empty rule base. The fuzzy rules can be henceforth extracted and removed during the training process based on a novelty of a new incoming training pattern and the contribution of an individual fuzzy rule to the system output. A single-pass manner prevails in the training process, thus keeping the virtual memory demand on a low-level and expediting model updates. A prominent aspect in PANFIS is the building of ellipsoids in arbitrary position (respecting local correlation between variables) connected with a new projection concept to form the antecedent parts in terms of linguistic terms (fuzzy sets). This is opposed to [15], where all operations are drawn directly in multidimensional level nonaxis parallel thus hardly interpretable for an expert/user. This also means that we are achieving fuzzy rules in a classical interpretable sense. The inference scheme is, however, still using the high-dimensional ellipsoidal representation in arbitrary directions as follows:

$$R_i = \exp(-(X - C_i)\Sigma_i^{-1}(X - C_i)^T) \tag{1}$$

where C is the center or template vector of ith rule, $C \in \Re^{1\times u}$, and X is an input vector of interest $X \in \Re^{1\times u}$. Σ_i is a dispersion or covariance matrix $\Sigma_i \in \Re^{u\times u}$ of ith rule, whose elements are the spreads of the (axis parallel) multidimensional Gaussians in each direction (dimension) σ_{ki} $k = 1, 2 \ldots, u$ and $i = 1, 2, \ldots, r$. The inference scheme is written as follows:

$$y = \sum_{i=1}^{r} w_i\phi_i = W\Psi = \frac{\sum_{i=1}^{r} R_i w_i}{\sum_{i=1}^{r} R_i}. \tag{2}$$

In this TS-type NFS, W labels the output parameters $W_i = k_{0i} + k_{1i}x_1 + \cdots + k_{ui}x_u$, $W \in \Re^{1\times(u+1)r}$, and $\Psi \in \Re^{(u+1)r\times 1}$ epitomize cluster or generalized (fuzzy) firing strength stemming from the firing strength of each fuzzy rules R_i.

Another novel aspect concerns the rule pruning methodology, which employs an extended rule significance (ERS) concept, supplying the blueprint of rule contributions. It is extended from the concept in SAFIS approach [13] by integrating hyperplanes (in lieu of constant) consequents and generalizing to ellipsoids in arbitrary position, allowing to prune the rules directly in the high-dimensional learning space. Two fuzzy sets, which are similar to each other, are grouped and are in turn blended to be one single fuzzy set exploiting a kerned-based metric [18] in conjunction with a transparent explanatory module (rule).

The rules are dynamically evolved based on a potential of the data point being learned by the use of datum significance (DS) criterion. The initialization of new rules is furthermore consummated in a new way assuring ε-completeness of the rule base as well as fuzzy partitions, thus leading to an adequate coverage of the input space. Whenever no new rule is evolved, the focal points and radii of the fuzzy rules are adjusted by means of extended self-organizing map (ESOM) theory to update neighboring rules with a higher intensity than rule lying farer away.

Another important facet of PANFIS is the use of enhanced recursive least square (ERLS) method (an extension of conventional recursive least square (RLS), which is widely used in the ENFSs community to adjust the fuzzy consequences), which has been formally proven to underpin the convergences of the system error and so the weight vectors being updated (a detail of this proof can be found in the appendix).

The remainder of this paper is organized as follows: Section II details the projection of arbitrary ellipsoids to fuzzy sets. Section III elaborates the incremental learning policy of PANFIS, which involves a rule base management. Section IV explores the empirical studies and discussions on numerous benchmark problems including artificial and real-world datasets. With these, the new evolving method will be compared against various other state-of-the-art evolving neuro-fuzzy approaches in terms of predictive quality and model complexity. Section V concludes this paper.

II. Projection of Ellipsoidal Clusters in Arbitrary Positions to Fuzzy Sets

The multidimensional membership function offers a convenient property in abstracting the data distribution in more natural way capable of triggerring ellipsoids in any directions in the feature space whose axes are not necessarily parallel. The underlying grievance is, however, to stipulate the fuzzy set representation of the hyperellipsoid in arbitrary positions entailing more faithful investigations as the covariance matrix constitutes a nondiagonal matrix thus allowing for more tractable rule semantics. The projection of ellipsoidal clusters in arbitrary positions to fuzzy sets is detailed herein to produce a classical-interpretable fuzzy rule (antecedent parts of rules can then be read as linguistic IF–THEN parts). The projection is undertaken after each incremental learning cycle, i.e., whenever a model update took place (see subsequent sections), thus representing the latest model version.

We advocate two possibilities of extraction, one faster but more inaccurate one, one slower but more accurate one, with the hope of landing on the fuzzy set representation of multivariate Gaussian function. Fig. 1 shows these two methods.

We suppose that C_i is the centroid of the ith fuzzy rule, whereas $\sum_{ij}^{-1}$ is the inverse covariance matrix where Σ_{ij} is the element in the ith row and jth column. Therefore, the fuzzy set is built as follows:

$$\upsilon_i = c_i \tag{3}$$

$$\sigma_i = \max_{k=1,..,u}\left(\frac{r}{\sqrt{\lambda_k}}\cos(\varphi(e_k, a_i)\right) \tag{4}$$

where λ_k is the eigenvalue with respect to kth dimension and e_k is for eigenvector for the corresponding kth dimension. μ_i is the modal value of the fuzzy sets (usually the center where the membership value is equal to 1) and σ_i is the spread of the set, according to a predefined α-cut; in our case, using Gaussian fuzzy sets, we apply an α-cut value of 0.6 whose cut reaches the inflection point, thus getting equal to the width sigma. Meanwhile, a_i epitomizes the vector representing ith axis $a_i = (0, 0, ..., 1, ...0)$, where the value of 1 implies the ith position. φ illustrates the angle spanned between a_i and e_k

$$\varphi(e_k, a_i) = \arccos\left(\left|\frac{e_j^T a_i}{|e_j|\,|a_i|}\right|\right). \tag{5}$$

The demerit of the first approach is computationally prohibitive especially in overcoming the high-dimensional input features as the eigenvalue of the inverse covariance matrix should be foreseen in every training episode. In lieu of the former, the second approach is put forward as an alternative

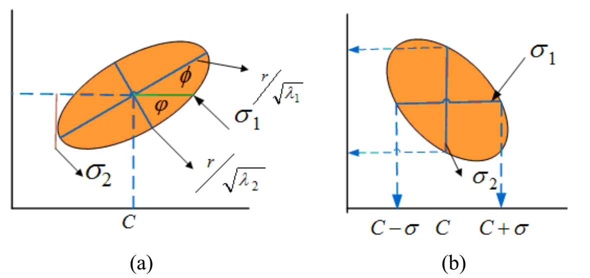

Fig. 1. Projection of the ellipsoids in arbitrary positions to fuzzy sets. (a) Eigenvalues and eigenvectors method. (b) Intersection or cutting point to ellipsoids method.

spotlight. The point of departure is the radii of the fuzzy sets can be granted by the distance from the center to the cutting point or the axis–parallel intersection with the ellipsoid, which is shown as follows:

$$\sigma_i = \frac{r}{\sqrt{\Sigma_{ii}}} \tag{6}$$

where Σ_{ii} is the diagonal element of the inverse covariance matrix. This method can expedite the training process, which is appealing in the realm of incremental learning, albeit this method can form too tiny spread in the case of large coverage span of the ellipsoid rotated $\sim 45°$.

III. Rule Base Management of PANFIS

All learning modules of PANFIS are outlined by this section where the principal constituents of PANFIS learning framework are specifically apportioned into five subsections as follows: A) rule updating with ESOM theory; B) recruitment of the fuzzy rules with DS concept; C) determination of the antecedent fuzzy rule based on ε-completeness criterion; D) pruning of inconsequential rules and merger of similar fuzzy set; and E) identification and adaptation of the consequent parameters. All of these are elaborated hereafter in the sequel.

A. Rule Updating With ESOM

Self-organizing map (SOM) theory is originated by Kohonen [26]. This method has prompted many researchers to embed it in their ENFSs to underpin their self-organizing property. The crux of the SOM method is appealing to progressively adjust the focal points (centers) of the Gaussian functions thus adapting to the changing patterns in the possible nonstationary streaming data. Nevertheless, a key grievance of this approach is that a spatial proximity in determining the winner is obtained via the Euclidean distance. That is, the distance is only calculated based on points without considering the interdependence between the input variables and the size of the fuzzy region. Apart from that, this method is unable to polish up the radii of the clusters. To correct this shortcoming, ESOM method resorts in overcoming these drawbacks with the compatibility measure of the firing strength and the use of sequential maximum-likelihood estimation proposed by [27] in adapting the zone of influence of the cluster. Hence, the center C_i and size of fuzzy region Σ_i of ith rule are updated by

$$C_i^n = C_i^{n-1} + \beta^n R_\upsilon^n h_\upsilon^n (X^n - C_i^{n-1}) \tag{7}$$

$$\Sigma_i^n = (N_i/N_i+1)\Sigma_i^{n-1} + (1/N_i+1)(X-C_i)^T(X-C_i) \tag{8}$$

$$h_i^n = \exp(-(C_\upsilon^n - C_i^n)^T \Sigma_i^{-1}(C_\upsilon^n - C_i^n)) \tag{9}$$

where β^n is the learning rate, R_υ^n is the firing strength of winning rule, and h_i^n is the neighborhood function.

$$\beta^n = 0.1\exp(-n). \tag{10}$$

The neighborhood function h_i^n estimates the similarities of two neighboring rules, whose value supplies a distance between the winner and the other rules. One may comprehend that more priority of adaptation is rewarded to a rule, which lies in a closer proximity to the winner supplying a larger value of the neighborhood function, whereas a more distant distance may be a factor of a less adjustment priority delivered setting less value of the neighborhood function as the updated rule is deemed dissimilar with the winner. Furthermore, the learning rate decays exponentially as more training episodes have been experienced by PANFIS assuming a more expedient rule base has been established in the end of training process.

B. Recruitment of Fuzzy Rules With DS

On the one hand, an ideal state to augment a fuzzy rule base is when a datum fed evokes a high system error. On the other hand, an extraneous fuzzy rule should be tailored when a model traverses an unmanageable region in the input space, which is untouchable with the existing fuzzy rules. These two prognoses hence ascertain a construction of a supplementary rule fostering the rule base from the presence of the next training samples located adjacent the newly created rule. In the DS concept amalgamated in PANFIS, the potential of the injected datum is designated by the statistical contribution of a datum when the number of observations approaches to infinity. For clarity, the approximation not only focuses the DS in the current condition, which is deemed susceptible with outlier, but also culminates up to the end of training process.

An algorithmic backbone of the DS criterion was proposed in [10] and [13] as an effective cursor to point out a high-potential datum managing troublesome data streams. Unfortunately, its default version is unable to be directly mounted in PANFIS learning engine as it is devised in the learning platform of hyperspherical rules and constant output parameters (whereas we are exploring hyperplane consequents and rules in arbitrary position in the learning space). Some amendments are to remedy these deficiencies undertaken in this paper, rendering the original version compatible with PANFIS learning platform. The adjustment process initiates with an introduction of the mathematical definition of the DS method as follows:

$$D_i = |e_n| \int_X \exp\left(-\frac{1}{k}\left(\frac{(X-X_n)\Sigma_i^{-1}(X-X_n)}{(X-C_i)\Sigma_i^{-1}(X-C_i)}\right)\frac{1}{S(X)}dx \tag{11}$$

where X_n is the latest datum extracted from input manifold $X_n \in \Re^u$, $S(X)$ is the range of the input X, and e_n is the system error denoting the degree of accuracy of PANFIS in the nth training observation, which is able to be written as follows:

$$|e_n| = |t_n - y_n| \tag{12}$$

y_n is the output of PANFIS in the nth episode, conversely, t_n is the target value in nth episode. By performing u-fold numerical integration, we may obtain as follows:

$$D_n = |e_n| \frac{\det(\sum_{i+1})^u}{S(X)}. \tag{13}$$

The significance of the new datum is defined as its statistical contribution to PANFIS's output. Vice versa, the premise part of PANFIS is a hyperellipsoid in arbitrary positions, whereby each membership function delineates the partition of the input space. Without loss of generality, the sampling data distribution $S(x)$ can be replaced by the total volumes of existing fuzzy rules. Therefore, (10) can be further written as follows:

$$D_n = |e_n| \frac{\det(\sum_{r+1})^u}{\left(\sum_{i=1}^{r+1} \det(\sum_i)\right)^u}. \tag{14}$$

If an observation complies with (11), the new datum suffixes to be hired as an extraneous rule. As $g < D_n$, the DS method appraises this datum high descriptive power and generalization potential. Conversely, if $g > D_n$, the current fuzzy rules have confirmed its completeness in seizing the available training stimuli. In this regard, the ESOM theory is taken place with the hope of refining the positions of the current fuzzy rules. The guideline in allocating the value of g is in addition illustrated elsewhere in this paper.

1) Initialization of New Fuzzy Rules Based on ε-Completeness Criterion: A selection of a new antecedent fuzzy set constitutes an indispensable constituent in the automatic proliferation of the fuzzy rules. It should be comprehensively formulated to achieve an adequate coverage in accommodating the universe discourses. In PANFIS learning platform, the allocation of the new premise parameters is enforced according to the profound concept of the ε-completeness criterion deliberated by [25]. In this viewpoint, for any inputs in the operating range, there exists at least a fuzzy rule so that a match degree is no less than ε for every individual input injected. Whenever the new fuzzy rule is appended as a reaction to troublesome training datum supplied, PANFIS shall plug in this input datum as the focal point or center of the new ellipsoidal rule. Furthermore, the size of the new ellipsoidal region is triggered by setting the width of the Gaussian function as follows.

Remark: In fuzzy applications, ε is usually selected as 0.5 [30].

$$C_{i+1} = X^n \tag{15}$$

$$\operatorname{diag}(\Sigma_{i+1}) = \frac{\max(|C_i - C_{i-1}|, |C_i - C_{i+1}|)}{\sqrt{\ln(\frac{1}{\varepsilon})}}. \tag{16}$$

One should envisage the width of multidimensional membership function playing prescriptive roles to PANFIS's final output because of forming a spread of the antecedent and a zone of influence of the ith fuzzy rule. For instance, too large value of the width for the Gaussian function influences to be averaging, too small value of it leads to be overfitting [29].

C. Pruning of Inconsequential Rules and Merger of Similar Fuzzy Sets

This section aims to explore the rule pruning and rule merging adornments of PANFIS. That is, these peculiar leverages are intended to attain an economical and interpretable rule base while sustaining its best predictive quality.

1) Pruning of Inconsequential Fuzzy Rules Based on ERS *Concept:* PANFIS inherits the rule base simplification technology of the so-called generalized growing and pruning radial basis function (GGAP-RBF), SAFIS [10], [13]. As foreshadowed, its predecessor is hampered to PANFIS algorithm as it is only compatible to the zero-order TSK fuzzy system and the unidimensional membership function environments. To suit it in PANFIS learning scenario, numerous customizations are made. Contribution of the ith rule to the overall system output for an input vector X_n is defined in a very similar way to the method for an online input sensitivity analysis and an input selection used in [34] by

$$E(i,n) = |\delta_i| \frac{R_i(x_n)}{\sum_{i=1}^{r} R_i(x_n)} = |\delta_i|\, E_i \tag{17}$$

$$E_i = \int_X \exp(-(X - C_i^n)^T \Sigma_i^{-1} (X - C_i^n)) \frac{1}{S(X)} dx \tag{18}$$

where $\delta_i = \sum_{k=1}^{u+1} w_{1i} + w_{2i}x_1 + \cdots + w_{ki}x_k + \cdots + w_{u+1,i}x_u$ outlines the output contribution of the ith fuzzy rule, whereas E_i denotes the input contribution of the ith fuzzy rules. Furthermore, the rule significance is defined as the statistical contribution of the fuzzy rule when the number of observation approaches to infinity $n \to \infty$. Referring to [10] and [13], the input contribution E_i can be derived by means of (18).

Generally speaking, the size of the inverse covariance matrix $\sum_i^{-1}$ is much less than the size of the range X. Hence, (18) can be simplified as follows:

$$E_i \approx \frac{1}{S(X)} (2 \int_0^\infty \exp(-(\frac{X^2}{\sum_i}))dx)^u \tag{19}$$

$$E_i = \frac{(2)^u}{S(X)} (\lim_{z\to\infty} \int_0^z \exp(-\frac{X^2}{\sum_i})dx)^u. \tag{20}$$

Via $u-$fold numerical integration for any arbitrary probability density functions $p(\mathbf{x})$ of the input data manifold $\boldsymbol{x}$ ($x \in \Re^u$), the solution of (20) can be expressed as follows:

$$\approx \frac{\pi^{u/2} \det(\sum_i)^u}{S(X)}. \tag{21}$$

On the one hand, the size of the range X namely $S(x)$ is not necessary to be computed. On the other hand, to extract the proper type of $S(x)$ is not trivial and time intensive. In fact, the sensitivity of the fuzzy rule is signified by the contribution of

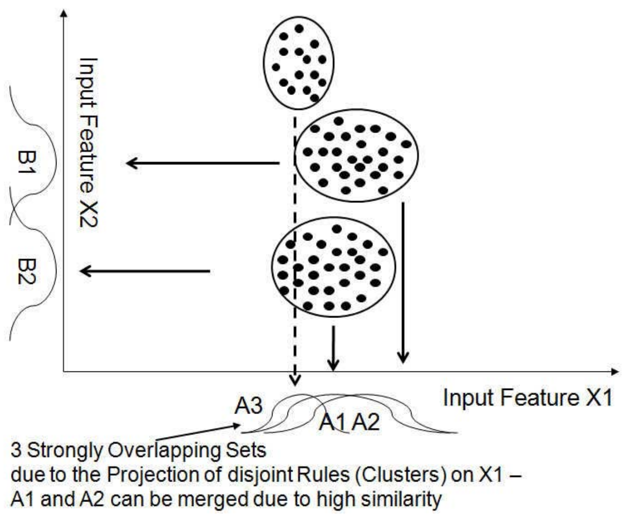

Fig. 2. Redundancy of two fuzzy sets.

the individual fuzzy rule over the total contributions of overall fuzzy rules. Therefore, we can justify the size of the range $S(x)$ to be replaced by the overall contributions of the fuzzy rules pinpointed by cluster volumes for the sake of flexibility. Accordingly, the statistical contribution estimation of the fuzzy rule is doable as follows:

$$E_{\inf}(i) = |\delta_i| \frac{\det(\sum_i)^u}{\sum_{i=1}^{r} \det(\sum_i)^u}. \tag{22}$$

If the contribution vector contains the fuzzy rule contributions less than equal k_{err}, then these fuzzy rules are classed as outdated fuzzy rules. They should be dispossessed to mitigate the rule base complexity thereby eradicating the vulnerabilities of overcomplex network structures gained. To the best of our knowledge, this method may excel most of other rule pruning mechanisms as they solely consider the rule contributions at the time an assessment is undertaken distorting the nature that a fuzzy rule may be a paramount ingredient to the system in the next training episodes inflicting unstable fluctuations of rule base size. This method nonetheless pays attention on the significance of consequent parameters demonstrating a particular operating region of a cluster as many of other variants relax it. Small consequent parameters incur little significance to the system.

2) Merger of Similar Fuzzy Sets: Although fuzzy sets are well scattered originally, they are still susceptible to be significantly overlapping to each other. It is undoubtedly inflicted by the projection concept, an example is shown in Fig. 2. If the membership functions are very similar to each other, they can be fused into one new membership function to obviate fuzzy rule redundancies and in turn to support an interpretable explanatory module (rule) of PANFIS. Similarity between two Gaussian membership functions A and B can be found by benefiting a kernel-based metric method comparing the centers and widths of two fuzzy sets in one joint formula [18] as follows:

$$S_{\mathrm{ker}}(A, B) = e^{-|c_A - c_B| - |\sigma_A - \sigma_B|} \tag{23}$$

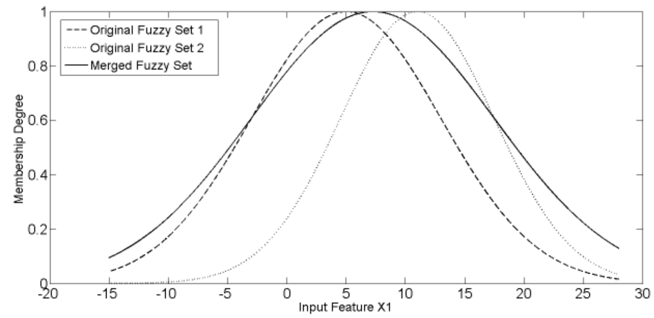

Fig. 3. Merging of two Gaussian fuzzy sets according to the α-cut.

having the following interesting properties:

$$S_{\mathrm{ker}}(A, B) = 1 \Leftrightarrow |c_A - c_B| + |\sigma_A - \sigma_B| = 0 \Leftrightarrow c_A = c_B \wedge \sigma_A = \sigma_B \tag{24}$$

$$S_{\mathrm{ker}}(A, B) < \varepsilon \Leftrightarrow |c_A - c_B| > \delta \vee |\sigma_A - \sigma_B| > \delta. \tag{25}$$

Another similarity measure as in [45] can be applied in lieu of kernel-based metric. The similarity measure in [45] is exploited to designate whether a new fuzzy set is tailored or the use of current fuzzy set is sufficient as a composition of new fuzzy rule. The first condition means that only in a case of a perfect match, the similarity measure between two fuzzy sets has the maximal degree of one. Conversely, the second condition assures that an embedded set is not coupled with a set covering it having a significantly larger width, which would result in a too inexact representation of the data in one local region. Two fuzzy sets are solely coalesced, if the similarity degree exceeds a tolerable value $S_{\mathrm{ker}} \geq 0.8$. The threshold is chosen as 0.8 as it is already delved in [18] as a coherent option to steer the tradeoff of dissimilarity and readability.

The merging itself is performed by the following formula [18]:

$$c_{\mathrm{new}} = (\max(U) + \min(U))/2 \tag{26}$$

$$\sigma_{\mathrm{new}} = (\max(U) - \min(U))/2 \tag{27}$$

where $U = \{c_A \pm \sigma_A, c_B \pm \sigma_B\}$, the underlying construct is to reduce the approximate merging of two Gaussian kernels to the exact merging of two of their α-cuts for a specific value of α. Here, we choose $\alpha = e^{-1/2} \approx 0.6$, which is the membership degree of the inflection points $c \pm \sigma$ of a Gaussian kernel. This also guarantees ε-completeness of the fuzzy partition as cuts the outer contours of the two sets at membership value 0.6, thus arriving at a larger coverage span than the original sets. An example is shown in Fig. 3. This concept can be easily generalized to arbitrary fuzzy sets employing the characteristic spread of the sets as σ, along a specific α-cut (usually in [0.4, 0.6]).

D. Derivations and Adaptations of Fuzzy Consequent Parameters

As outlined in Section I, PANFIS plugs in ERLS method in deriving and updating the fuzzy consequent parameters, which feature a synergy between a local learning approach [11] and an extended RLS. Suppose we arrive at $n = z$ training episode and we are interested at minimizing the locally weighted cost function, so that the desired output might be perfectly replicated by PANFIS output as follows:

$$J_L = \sum_{n=1}^{z} (Y - X^T w_i)^T \Delta_i (Y - X^T w_i) \tag{28}$$

where Δ_i is a diagonal matrix, where its main diagonal elements comprise $\phi_i(x_k)$. Meanwhile, X is formed by $x_e = [1, x_1, ..., x_u] X \in \Re^{z\times(u+1)}$. Thereafter, (28) can be solved assuming the linear subsystems are loosely coupled with level of interactions expressed by ϕ_i.

$$J_L = \sum_{n=1}^{z} J_{Li}. \tag{29}$$

Finally, the optimum solution of the local subsystem w_i that minimizes the locally weighted cost function is obtained as follows:

$$w_i = (X^T \Delta_i X)^{-1} X^T \Delta_i T = \Psi^\wedge T \tag{30}$$

where $\Psi^\wedge$is the Moore–Penrose generalized inverse of the Ψ, the weight vector w_i is tangible to be unique when $\Psi^T\Psi$ is a nonsingular matrix. Therefore, there are various ways to compute the Moore–Penrose generalized inverse disclosed by publications [31], involving orthogonal projection method, orthogonal method, iterative method, and singular value decomposition [32] method. On the one hand, if $\Psi^T\Psi$ is a nonsingular matrix, the pseudoinversion or orthogonal projection is ideally applicable. On the other hand, if $\Psi^T\Psi$ tends to be singular, we use Tichonov regularization

$$\Psi^T\Psi = (\Psi^T\Psi + \alpha I) = X^T \Delta_i X = (X^T \Delta_i X + \alpha I) \tag{31}$$

where we elicit the regularization parameterα by [39]

$$\alpha \approx \frac{2\lambda_{\max}}{\text{threshold}} \tag{32}$$

where threshold set to 10^15 and $\lambda_{\max}$is the largest eigenvalue. After the weight matrix Whas been optimally assembled by the local approach, it is allowed to be polished up recursively on the fly by the ERLS method. To detail this concept, the point of departure is to assure the convergence of the system error and weight vector being adapted.

To obviate the convergence issue in RLS deliberated in [33], the RLS can be extended merely inserting an additional constant α conferring a noteworthy effect to foster the asymptotic convergence of the system error and weight vector being adapted, which acts like a binary function. As reciprocal impact, the consolidated constant α shall be one if the approximation error $\hat{e}$ is bigger than the system error e. Otherwise, it shall stay constant at zero implying the preference of the existing weight vector rather than adjusting it. In other words, the constant α is in charge to regulate the current belief of the weight vector W. It can be presented mathematically as follows:

$$\alpha = \begin{cases} 1, \left|\widehat{e}\right| \geq |e| \\ 0, \text{otherwise} \end{cases} \tag{33}$$

$$L(n) = Q_i(n-1)x_e(n)(\Psi(n)^{-1} + x_e^T(n)Q_i(n-1)x_e(n))^{-1} \tag{34}$$

TABLE I

SENSITIVITY OF k_{err}, g

	$g = 0.1$	$g = 0.01$	$g = 0.001$
$k_{err} = 0.1$	(2,0.265)	(2,0.255)	(2,0.258)
$k_{err} = 0.01$	(3,0.245)	(3,0.247)	(3,0.252)
$k_{err} = 0.001$	(5,0.242)	(4,0.238)	(4,0.241)

$$Q_i(n) = (I - \alpha L(n-1)x_e^T(n))Q_i(n-1) \tag{35}$$

$$w_i(n) = w_i(n-1) + \alpha L(n)(t(n) - x_e^T(n)w_i(n-1)) \tag{36}$$

where we assign $w_{i0} = 0$ and $Q_{i0} = \omega I$. Note that the system error is defined by (12), whereas the approximation error $\hat{e}$ is the distortions between the true output values and the predictions of PANFIS while modeling one step ahead of the system behavior formulated as follows:

$$\hat{e} = |t_n - \Psi_n W_{n-1}|. \tag{37}$$

We ought to comprehend that when the rule growing module is turned on, the covariant matrix $Q_{r+1}$is set as $Q_{r+1} = \omega I$. Regarding to the above equations, they imply that the adaptations of the weight vector W are disallowed when the approximation error $\hat{e}$ is less than the system error e. The consummation of the RLS method substantiates PANFIS's predictive quality, as it aids PANFIS to always strike a reasonable tradeoff of the system error and weight vector convergences. The proofs are given in the appendix. Moreover, we also compare PANFIS's performance employing ERLS and default RLS in the numerical examples. Generally speaking, the ERLS boosts the predictive fidelity of PANFIS when it predicts the footprints of the system.

E. Sensitivity Analysis of the Preset Parameters

This section purposes to confer the sensitivity of PANFIS's predefined parameters. The coherent way in setting the preset parameters is explored to render PANFIS more ergonomic. Accordingly, the Box–Jenkins gas furnace dataset disseminated in [46] is cast to probe the sensitivity of k_{err}, g in PANFIS's algorithms. This dataset is compiled 290 input/output data points where its input attributes are process input variable $u(k)$, CO_2 concentration in off gas. Conversely, the process output $y(k)$ denotes the output attribute. More specifically, the overall 290 training stimuli feed PANFIS and the nonlinear dependence of the system is orchestrated as follows:

$$y(t) = f(y(t-1), u(k-4)). \tag{38}$$

In the following, the parameters k_{err}, $g$are opted as 0.1, 0.01, and 0.001. Table I encapsulates the experimental results in terms of the number of rules and nondimensional error index (NDEI).

Arguably, the threshold g is a noteworthy parameter in controlling whether PANFIS augments its structure (stability), or tunes the current structure (plasticity). Hence, the threshold g fully steers a stability–plasticity dilemma disseminated by [28], which is a cursor to notify an action to be carried out in the rehearsal process. On the one hand, one may comprehend that a lower value of g chosen would induce a better predictive accuracy of PANFIS. It, however, always lands on a more complex fuzzy rule base invoking more fuzzy rules to be crafted thereby suppressing the structural efficiency and transparency and vice versa. On the other hand, the lower value of k_{err} yields the more frugal rule base and the higher value of it evokes more demanding rule base assembled and in turn solidifies the predictive quality of PANFIS. PANFIS is in essence supposed to gain a balance between them. To actualize a synergy between the predictive accuracy and the compactness of the rule base, we fix $k_{\text{err}} = 10\%g$, $g = 10^{-u}$ usually acquiring good results with respect to the numerical results in Table I.

IV. Numerical Examples

To evaluate the efficacy of PANFIS learning policy, three empirical studies in the real-world and synthetic streaming data are carried out to evaluate the self-reorganizing and self-correcting mechanisms of PANFIS. The streaming data employed herein are: standard and poor (S&P) index data, hyperplane data, and NOx emission in a car engine data. PANFIS is benchmarked with its counterparts in the ENFS to obtain profound insight of PANFIS's efficacy in which the qualities of the benchmarked algorithms are assessed by virtue of root-mean-square error (RMSE), and NDEI written in the compact mathematical forms as follows:

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N}(t(n)-y(n))^2}{\sqrt{N}}}, \quad NDEI = \frac{RMSE}{STD(t)}. \quad (39)$$

A. Prediction of S&P Index Time Series

This section analyzes the sustainable self-reorganizing behaviors of PANFIS in addressing the S&P-500 market index dataset. The dataset encompasses 60-year daily index value, which can be found in the Yahoo! Finance website. A total of 14 893 data were acquired from January 3, 1950 to March 12, 2009. Importantly, this dataset is frequently unable to be modeled with the traditional NFSs, which have a prefixed structure, because of the nonlinear, erratic and time-variant behaviors of the underlined dataset. All the training data embark to PANFIS algorithm to pursue the self-correcting mechanism of PANFIS in recognizing the underlying characteristics of the S&P-500 index time series dataset. PANFIS is benchmarked against its counterparts such as DENFIS, eTS, simp_eTS, adaptive network based on fuzzy inference system (ANFIS), linear regression (LR), evolving fuzzy neural network (EFuFNN), and self organizing fuzzy associative machine (SeroFAM) [11], [29], [34]–[36], [47] so as to benchmark the superiority of PANFIS algorithm contrasted with the state-of-the-art works. Apart from that, the ERLS and RLS methods are contrasted to abstract the merit of the ERLS against RLS methods and in turn to illustrate the effect of the ERLS method in fortifying PANFIS's performance. Table II tabulates the consolidated results of all benchmarked systems.

The input and output relationship of the system is governed by the following equations:

$$y(t+1) = f(y(t-4), y(t-3), y(t-2), y(t-1), y(t)). \quad (40)$$

TABLE II
Forecasting 60 Years of S&P-500

MODEL	TOTAL PARAMETER	RULE BASE PARAMETER	NUM OF SETS	NUM OF RULES	NDEI
PANFIS-ERLS	144	144	3-4-3-4	4	0.014
PANFIS-RLS	144	144	3-4-3-4	4	0.0142
LR	14899	6	-	-	0.157
DENFIS	126	126	-	6	0.02
ANFIS	320	15213	-	32	0.015
EFuNN	1143	1143	-	114.3	0.154
SeroFAM	290	290	-	29	0.027
eTS	75	75	-	14	0.015
Simp_eTS	39	39	-	7	0.045

Note that the target value is the absolute value of the S&P-500 daily index. Fig. 4(a) pictorially shows the plot of one-step-ahead outputs of PANFIS versus the target values of the training samples. Fig. 4(b) shows the same as Fig. 4(a), however, it merely portrays the last 200 samples. Fig. 4(c) shows the deviations between the output predictions and the true output values occurring in the rehearsal process. Fig. 4(d) shows the trace of PANFIS rules while attempting to learn all regimes of the training data.

From Table II, we may deduce that PANFIS showcases a favorable performance, which outperforms the other approaches conferring the highest modeling accuracy while employing the most compact and parsimonious structure. Surprisingly, the ANFIS and LR methods, which are subsumed as the nonevolving algorithms, can marginalize or can be comparable the DENFIS, simp_eTS, EFuNN, and SeroFAM predictive accuracy. It is worth stressing that these methods necessitate a complete dataset at hand *a priori* enduring a retraining phase benefiting from an up-to-date dataset whenever they gather new training stimuli and imposing the explosion of number of parameters stored in memory.

Note that the total parameter shown in Table II encapsulates the number of rule base parameters and training data made use to model updates whereas the rule base parameter implies the parameters invoked to form fuzzy rules mainly affected by number of input dimensions and fuzzy rules. Conversely, the evolving method is remarkable favoring to learn any data points once in time without underpinned by the expert knowledge. One may argue no substantial deterioration of LR's predictive accuracy when smaller number of training data is maintained in the memory. LR method is, even so, incompetent to commence its learning process with single training datum thus disallowing viability in truly online situation usually only small snapshot of training data available at hand before the process runs.

Generally speaking, SeroFAM and EFuNN are inferior to other algorithms. Obviously, this condition is stimulated as both approaches adopt Mamdani-based fuzzy system, whose premise and consequent parameters comprise the membership functions. It is tangible that the TSK fuzzy system employs precise mathematical models in the consequent part, which

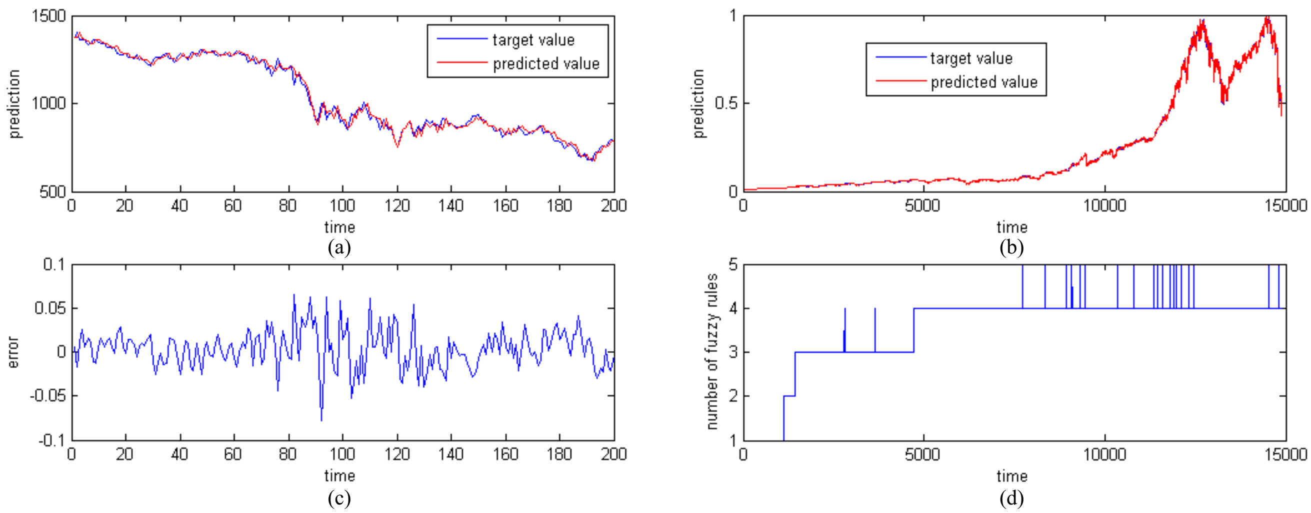

Fig. 4. S&P-500 dataset. (a) and (b) Modeling of the S&P-500. (c) System error. (d) Fuzzy rule evaluation in one of trials.

arguably incur higher predictive accuracy. Nevertheless, the trait of Mamdani-based fuzzy system is fairly endorsed, which yields more transparent rule semantics owing to endued by the fully linguistic rules.

For clarity, the evolving layer in the EFuFNN solely self-reorganizes [48] imposing the worst performance. An in-depth look of multidimensional membership function may bear to additional parameters to dwell in the memory. This type can, nonetheless, trigger a more natural representation of ellipsoidal clusters whose axes are not necessarily parallel to input axes. One may envision the number of parameters hinging to the number of rules bred where the number of network parameters of PANFIS is still manageable in this paper case, notwithstanding simp_eTS consumes a smaller number of parameters than PANFIS. It is worth noting that simp_eTS devised with unidimensional membership function flourishing spherical clusters.

It can be seen in Fig. 4 that PANFIS is capable of swiftly addressing the nonstationary component of the training data automatically recruiting and evicting fuzzy rules in a timely fashion. A trivial analogy is the evolution of individuals in nature during life cycle, specifically the autonomous mental development of humans, which usually starts from scratch without any knowledge and important information can be memorized in a form of linguistic rules afterward. Vice versa, outdated knowledge can be forgotten without substantially affecting their developments. More interestingly, PANFIS mimics the true output values closely in which the system error hovers around quiet small values. In the following, the ERLS method navigates to a substantial improvement for PANFIS predictive fidelity in which it weakens when PANFIS exploits the standard RLS method. Indeed, the rule merging strategy of PANFIS is beneficial to diminish the number of fuzzy sets and to warrant the transparency of the rule base. In contrast, other algorithms except SeroFAM in this circumstance exclude this component.

B. Hyperplane Data Streams

This paper case alludes the viability of PANFIS in the classification problem using artificially generated data streams by means of synthetic stream generator from the massive online analysis [44]. In a nutshell, this task embodies a binary classification problem laying out a random hyperplane in d-dimensional Euclidean space as a decision boundary. In the following, the point $x_i = (x_1, \ldots, x_d)$ is placed in the hyperplane, if it is in line with the following equation:

$$\sum_{i=1}^{d} w_i x_i = w_0. \tag{41}$$

In general, the hyperplane can establish a binary classification problem in which $\sum_{i=1}^{d} w_i x_i > w_0$ pinpoints a positive class, whereas $\sum_{i=1}^{d} w_i x_i < w_0$ approbates a negative class. The landmark of this dataset interestingly manifests a drift in which the underlying characteristic is to blend two pure distributions in probabilistic fashion. It can be delved that the data are emanated by the first distribution with the probability of one at the beginning. Henceforth, this probability attenuates and in turn lands on the second distribution. In connection with the experimentation, the data streams entangle in sum 120k points where the drift is asserted after 40k samples.

In this viewpoint, we accomplish a periodic hold-out test to figure out the self-reorganizing property of PANFIS swiftly coping with the drift in the streaming examples. That is, the first 1200 data samples are embarked in the first experiment in which 1000 data points are enumerated to training process and 200 data are planned to a validation phase. In the second experiment, the next 1200 data are cast and the portion of training and testing data are tantamount to the first experiment, and so on. Our brand new algorithm is benchmarked with its counterparts such as ANFIS [47], eTS [11], simp_eTS [29], and FLEXFIS+ [18]. Fig. 5 shows the trace of the classification rate, the rule evolution in the 41st experiment and the number of fuzzy rules scattered in each hold-out test. Meanwhile, Table III summarizes the consolidated results of the benchmarked system.

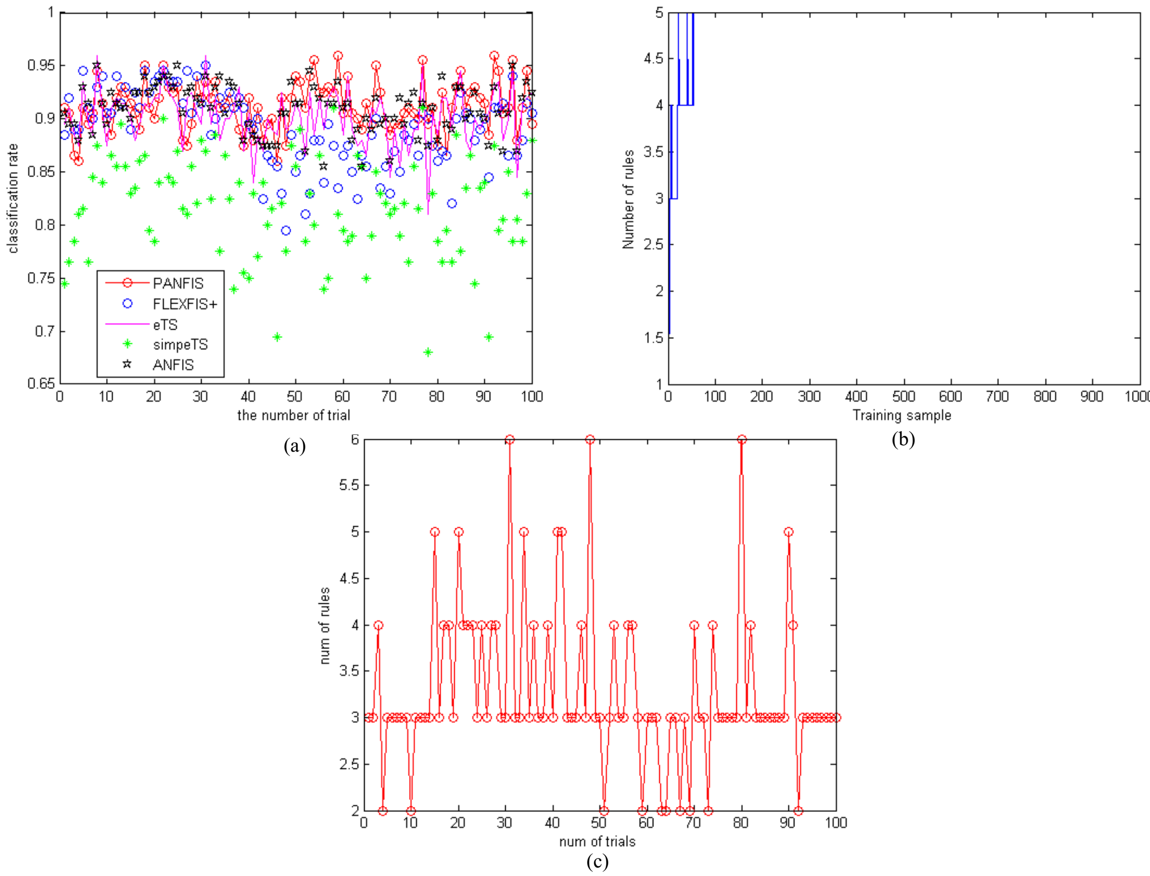

Fig. 5. Hyperplane dataset. (a) Trace of the classification rate in each trial. (b) Fuzzy rule evolution in one of trials. (c) Fuzzy rule evolution of whole process.

TABLE III
MODELING OF HYPER-PLANE

MODEL	TOTAL PARAM ETER	RULE BASE PARAMETE R	NUM OF RULE S	CLASSIFI CATION RATE
PANFIS-ERLS	114.84	114.84	3.19	0.9155
PANFIS-RLS	114.84	114.84	3.19	0.9126
ANFIS	120.256	256	16	0.909
eTS	164.21	164.21	14.49	0.9016
Simp_eT S	62.53	62.53	5.23	0.8171
FLEXFI S+	115.52	115.52	7.22	0.8906

From Table III, one may appraise PANFIS sheds a fruitful impetus in learning in the nonstationary environment context in which PANFIS provides more convincing performance than FLEXFIS+, eTS, and simp_eTS. It is worth noting that PANFIS holds the best predictive accuracy and salvages the rule base cost. More specifically, Fig. 5(b) shows the automatic rule growing and pruning of PANFIS, which is effectual to prevail drifts contained by the system without the detrimental costs to the model evolution. The rule pruning mechanisms, in the following, are triggered in the early episodes of training process. PANFIS commences its training process with the first training datum and the first fuzzy rule is crafted by virtue of this datum. This rule probably does not contribute significantly during its lifespan, it is therefore a subject of rule pruning module. The assembled rule base is not sufficiently mature in the early training episodes thus leading to high system errors in learning the training samples and breeding the extraneous fuzzy rules. These rules, are later on inactive, the rule pruning module is hence activated to evict these rules. Fig. 5(b) shows the 41st periodic hold-out process where the drift commences to ensue. That is, the changing data distribution evokes that the previously appended rules are invalid to delineate the current data trend. As more training stimuli are injected, the rule base evolution ends up more stable rule evolution where the rule growing and pruning do not ensue too frequently. That is, the refinements of input and output parameters are merely carried out as the current rule base is already representative to capture the training data.

An in-depth look elicits that the performance of PANFIS worsens in the 41st trial owing to drift. As the rule base is refurbished in the next training episode, the performance of PANFIS is boosted, which is in line with the increase of the classification rates in the next experiments. By extension, Fig. 5 also shows the overview of number of fuzzy rules in every trial, which endorses the flexibility of PANFIS learning engine as it is capable of orchestrating its rule base size according to the knowledge recognized. It is worth stressing that ANFIS is inherently an offline method and depletes much training overhead emphasized by considerable parameters

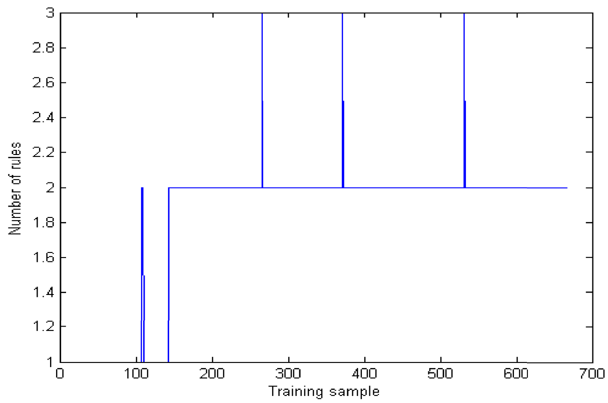

Fig. 6. Trace of rule generation.

number saved in the memory, notwithstanding it excels the simp_eTS and FLEXFIS+, which are fully online algorithms. In contrast, the simp_eTS manipulates unidimensional membership functions, albeit it deposits the smaller number of parameters than that PANFIS.

C. Modeling of NOx Emission on Car Engine Tests

The last study case is to model a real-world dataset driven from tests with a real car [41]. Interestingly, the car engine data implicitly comprise various process changes in form of different operation modes, occurring in a wide range of the engine operation map. In particular, the two essential main influencing factors for engine control, namely rotation speed and torque were not kept in a specific driving mode range (i.e., constantly to simulate driving on a motor highway), but from time to time varied for simulating different driving behaviors. This brings in some sort of changing dynamics of the car engine behavior, also with respect to NOx emissions and thus is expected to be modeled by static models (not being able to adapt to new system states) with insufficient accuracy. We will also verify this when comparing our novel method with nonevolving fuzzy systems/models.

Nonetheless, the main task of PANFIS in this problem is to predict NOx emissions from the exhaust in a car engine by virtue of four input attributes as follows: N: engine rotation speed (rpm). $P2$: pressure offset in cylinders (bar). T: engine output torque (Nm). Nd: speed of dynamometer (rpm). The nonlinear dependence of the system is regulated by the following equation:

$$\text{No}x_n = [N_{n-4}, P2_{n-5}, T_{n-5}, Nd_{n-5}]. \tag{42}$$

This dataset comprises 826 data pairs in which 667 data are captured for the training purpose, whereas 159 of which are injected to test the generalizing ability of the produced rule base. By extension, PANFIS is benchmarked against its counterparts such as eTS, Simp_eTS, ANFIS, FAOS-PFNN, and bayesian art fuzzy inference system [11], [12], [29], [37], [38], [47] in which their predictive accuracies are assessed by the RMSE. Table IV tabulates the consolidated results of all benchmarked system in the withheld evaluation set of 159 data points whereas Fig. 6 displays the evolution of fuzzy rules in the training phase.

TABLE IV
MODELING OF NOx EMISSIONS DATASET

MODEL	TOTAL PARAMETER	RULE BASE PARAMETER	NUM OF SET	NUM OF RULES	NDEI
PANFIS-ERLS	36	36	2-1	2	0.578
PANFIS-RLS	36	26	2-1	2	0.607
ANFIS	1034	208	-	16	0.626
eTS	64	64	-	6	1.301
Simp_eTS	49	49	-	5	1.34
FAOS-PFNN	952	126	-	21	0.781
BARTFIS	247	247	-	19	0.607
FLEXFIS	70	70	-	5	0.589

From Table IV, PANFIS showcases the best result in which PANFIS can accurately resemble the footprints of 159 testing data, while swiftly retaining the most frugal rule base. On the one side, FAOS-PFNN and ANFIS, which are not devised to an online learning purpose, can land on the competitive modeling quality as PANFIS. Nevertheless, we contend oversized rule base yielded by these methods due to the absence of forgetting mechanism in inconsequential fuzzy rules. On the other side, FLEXFIS strikes the second best quality in terms of the structural load and in prognosticating the dynamic of NOx emission. One may discern that FLEXFIS undermines the logic of online learning scenario because of a compulsory pretraining process with the use of some recorded data.

V. CONCLUSION

This paper explores an agile ENFS termed as PANFIS. The efficacy and viability of PANFIS prototype have been exemplified through three study cases in miscellaneous benchmark problems. Nonetheless, the comparisons against numerous state-of-the-art works have been enforced, which astonishingly infers PANFIS as an overwhelming breakthrough to the evolving neuro-fuzzy field marginalizing its counterparts predictive accuracy and structural complexity facets. As a future work, the application of PANFIS in dealing with the classification case is the subject of our further investigation.

VI. APPENDIX

1) *Proof of System Error Convergence*

Theorem 1: If a nonlinear system is stable and it is desired to be modeled by benefiting PANFIS, it will always deliver a small finite value of the system error [6].

Proof: Assuming that the approximation error $\hat{e}$ converges at $n \geq n_0$, the approximation error always complies with both of two constrains [7], [46] $\lim_{n\to\infty} \left|\hat{e}\right| = 0$, $\max\left|\hat{e}(n)\right| \leq \delta$ with the use of ERLS method, where δ is an arbitrary small constant. The system error is defined in (12). Conversely, the approximation error is written in (39). The absolute value of the system error is illustrated as follows:

$$\begin{aligned} |e_n| &= |t_n - W_n\Psi_n| && \text{(A1)}\\ &= \left|I - L_n^T x_n\right| |t_n - W_{n-1}\Psi_n| && \text{(A2)}\\ &= \left|\frac{\hat{e}_n}{1 + \Psi_n x_e^T Q_{i,n-1} x_e}\right| \leq \delta. && \text{(A3)}\end{aligned}$$

It should be emphasized that if $\left|1+\Psi_n x_e^T Q_{i,n-1} x_e\right|$ is bounded and is not equal with -1 for throughout training process, it can be derived as follows:

$$=|e_n|\left|1+\Psi_n x_e^T Q_{i,n-1} x_e\right| = \left|\hat{e}_n\right| \leq \left|1+\Psi_n x_e^T Q_{i,n-1} x_e\right| \delta. \tag{A4}$$

To guarantee that (35) is encountered during the training process, it leads to:

$$1 \leq \left|1+\Psi_n x_e^T Q_{i,n-1} x_e\right| \leq S \tag{A5}$$

where S is a positive constant, because of the arbitrary small value δ, $\left|1+\Psi_n x_e^T Q_{i,n-1} x_e\right| \delta$ is certainly an arbitrary small value. Therefore, we acquire that the approximation error is convergent:

$$\lim_{n\to\infty} \left|\hat{e}\right| \cong 0. \tag{A6}$$

Herewith, the Theorem 1 is proven.

1) *Proof of Parameter Convergence*

Theorem 2: If a nonlinear system modeled by PANFIS is stable, the weight vector of PANFIS adapted by the ERLS approach will be bounded to a finite vector as time approaches infinity [7].

Proof: Using l_1 matrix norm, we consider two consecutive time instants $t$and $t \ - 1$, which are detailed as follows:

$$\begin{aligned}\left\|w_{i,n}-w_{i,n-1}\right\|_1 &= \left\|L_n\left|t_n - x_{e,n}^T w_{i,n-1}\right|\right\|_1 \\ &= \left\|Q_{i,n} x_{e,n}\hat{e}_n\right\|_1 \leq \left\|Q_{i,n}\right\|_1 \left\|x_{e,n}\right\|_1 \left\|\hat{e}_n\right\|_1\end{aligned} \tag{B1}$$

where

$$\left\|Q_{i,n}\right\|_1 \leq \sqrt{M}\left\|Q_{i,n}\right\|_2 \sqrt{\rho(Q_{i,n}^H Q_{i,n})} = \sqrt{M}\lambda_{\max}(Q_{i,n}) \tag{B2}$$

$\lambda_{\max}(Q_{i,n})$ is the maximum eigenvalue of the Hermittian matrix $Q_{i,n}$. From (34), $Q_{i,n}P_n$may be presented as follows:

$$Q_{i,n}x_{e,n} = Q_{i,n-1}x_{e,n}(\Psi_n^{-1} + x_{e,n}^T Q_{i,n} x_{e,n})^{-1}. \tag{B3}$$

As foreshadowed in Appendix A, $1 \leq \left|1+\Psi_n x_{e,n}^T Q_{n-1} x_{e,n}\right| \leq S$ thereby leading to:

$$= Q_{i,n-1}x_{e,n}(\Psi_n^{-1} + x_{e,n}^T Q_{i,n} x_{e,n})^{-1} \leq Q_{i,n-1}x_{e,n}. \tag{B4}$$

After that, Q_n may be written as follows:

$$\left\|Q_{i,n}\right\| = \left\|\frac{Q_{i,n}x_{e,n}}{x_{e,n}}\right\| \leq \left\|\frac{Q_{i,n-1}x_{e,n}}{x_{e,n}}\right\|. \tag{B5}$$

Thus, Q_n always complies with the following condition:

$$\left\|Q_{i,n}\right\| \leq \left\|Q_{i,n-1}\right\| \leq \ldots \leq \left\|Q_{i,1}\right\|. \tag{B6}$$

By a utilization of (B2), Q_n can be formed as a function of the largest eigenvalue as follows:

$$\sqrt{M}\lambda_{\max}Q_{i,n} \leq \sqrt{M}\lambda_{\max}Q_{i,n-1} \leq \ldots \leq \sqrt{M}\lambda_{\max}Q_{i,1}. \tag{B7}$$

We suppose that, at time t, the upper bound of x_e is b.

$$\sum_{k=1}^{u+1} x_{e,k} = b. \tag{B8}$$

Finally, we are able to express (B2) as the following inequality:

$$\left\|W_n - W_{n-1}\right\|_1 \leq \sqrt{M}\lambda_{\max}Q_1 b\left|\hat{e}_n\right|. \tag{B9}$$

As highlighted in (A6) $\lim_{n\to\infty}\left|\hat{e}\right| = 0$. Hence, it follows that the subtraction of the weight vector in the two consecutive time instants also approaches to zero as the time reach infinity or is bounded to a small value.

$$\lim_{n\to\infty}\left\|W_n - W_{n-1}\right\| \cong 0. \tag{B10}$$

Hence, we may infer that Theorem 2 is proven.

Acknowledgment

This paper reflects only the authors' views.

References

[1] L. A. Zadeh, "Soft computing and fuzzy logic," *IEEE Softw.*, vol. 11, no. 6, pp. 48–56, Nov. 1994.

[2] L. T. Whye and Q. Chai, "eFSM-A novel online neural fuzzy semantic memory model," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 136–157, Jan. 2010.

[3] E. Lughofer, *Evolving Fuzzy Systems—Methodologies, Advanced Concepts and Applications*. New York, NY, USA: Springer-Verlag, 2011

[4] W. A. Farag, V. H. Quintana, and G. Lambert-Torres, "A genetic-based neuro-fuzzy approach for modeling and control of dynamical systems," *IEEE Trans. Neural Netw.*, vol. 9, no. 5, pp. 756–767, Sep. 1998.

[5] E. Lughofer and P. Angelov, "Handling drifts and shifts in on-line data streams with evolving fuzzy systems," *Appl. Soft Comput. (Elsevier)*, vol. 11, no. 2, pp. 2057–2068, 2011.

[6] S. Wu and M. J. Er, "Dynamic fuzzy neural networks-a novel approach to function approximation," *IEEE Trans. Syst. Man Cybern., Part B, Cybern.*, vol. 30, no. 2, pp. 358–364, Apr. 2000.

[7] G. Leng, G. Prasad, and T. M. McGinnity, "An on-line algorithm for creating self-organizing fuzzy neural networks," *Neural Netw.*, vol. 17, no. 10, pp. 1477–1493, 2004.

[8] R. M. French, "Catastrophic forgetting in connectionist networks," in *Encyclopedia of Cognitive Science*, vol. 1. L. Nadel, Ed. London, U.K.: Nature Publishing Group, 2003, pp. 431–435.

[9] E. Lughofer and P. Angelov, "Handling drifts and shifts in on-line data streams with evolving fuzzy systems," *Appl. Soft Comput. (Elsevier)*, vol. 11, no. 2, pp. 2057–2068, 2011.

[10] G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 57–67, Feb. 2005.

[11] P. Angelov and D. Filev, "An approach to online identification of Takagi-Sugeno fuzzy models," *IEEE Trans. Syst., Man, Cybern., Part B, Cybern.*, vol. 34, no. 1, pp. 484–498, Feb. 2004.

[12] E. Lughofer, "FLEXFIS: A robust incremental learning approach for evolving Takagi-Sugeno fuzzy models," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1393–1410, Dec. 2008.

[13] H. J. Rong, N. Sundararajan, G. B. Huang, and P. Saratchandran, "Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and time series prediction," *Fuzzy Sets Syst.*, vol. 157, no. 9, pp. 1260–1275, 2006.

[14] J. A. Dickerson and B. Kosko, "Fuzzy function approximations with ellipsoidal rules," *IEEE Trans. Syst., Man, Cybern. Part B, Cybern.*, vol. 26, no. 4, pp. 542–560, Aug. 1996.

[15] A. Lemos, W. Caminhas, and F. Gomide, "Multivariable Gaussian evolving fuzzy modeling system," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 91–104, Feb. 2011.

[16] P. Angelov, "Evolving Takagi-Sugeno fuzzy systems from data streams (eTS+)," in *Evolving Intelligent Systems: Methodology and Applications*, P. Angelov, D. Filev, and N. Kasabov Eds., New York, NY, USA: Willey, Apr. 2010, pp. 21–50.

[17] M. Setnes, "Simplification and reduction of fuzzy rules," *Interpretability Issues In Fuzzy Modeling. Studies in Fuzziness and Soft Computing*, J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, Eds. vol. 128. New York, NY, USA: Springer-Verlag, 2003, pp. 278–302.

[18] E. Lughofer, J.-L. Bouchot, and A. Shaker, "On-line elimination of local redundancies in evolving fuzzy systems," *Evolving Syst.*, vol. 2, no. 3, pp. 380–387, 2011.

[19] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with fuzzy logic controller," *Int. J. Man-Mach. Stud.*, vol. 7, no. 1, pp. 1–13, 1975.
[20] W. Pedrycs, "An identification algorithm in fuzzy relational system," *Fuzzy Sets Syst.*, vol. 13, no. 2, pp. 153–167, 1984.
[21] P. Angelov and R. Yager, "A new type of simplified fuzzy rule-based systems," *Int. J. General Syst.*, vol. 41, no. 2, pp. 163–185, 2011.
[22] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its appfications to modeling and control," *IEEE Trans. Syst. Man Cybern.*, vol. 15, no. 1, pp. 116–132, Jan./Feb. 1985.
[23] J. L. Castro and M. Delgado, "Fuzzy systems with defuzzification are universal approximators," *IEEE Trans. Syst., Man Cybern., Part B, Cybern.*, vol. 26, no. 1, pp. 149–152, Feb. 1996.
[24] L. Wang, "Fuzzy systems are universal approximators," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Mar. 1992, pp. 1163–1170.
[25] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller. II," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 2, pp. 404–436, Mar./Apr. 1990.
[26] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.
[27] B. Vigdor and B. Lerner, "The Bayesian ARTMAP," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1628–1644, Nov. 2007.
[28] S. Grossberg, *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control*. Boston, MA, USA: Reidel, 1982.
[29] P. Angelov and D. Filev, "A simplified method for learning evolving Takagi-Sugeno fuzzy models," in *Proc. IEEE Int. Conf. Fuzzy Syst. FUZZ*, May 2005, pp. 1068–1073.
[30] S. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, pp. 267–278, 1994.
[31] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*. New York, NY, USA: Wiley, 1971.
[32] G. Huang, Q. Zhu, and C. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, Dec. 2006.
[33] K. J. Astrom and B. Wittenmark, *Adaptive Control*, 2nd ed. Reading, MA, USA: Addison Wesley, 1995.
[34] N. Kasabov and Q. Song, "DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time series prediction," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 144–154, Apr. 2002.
[35] N. K. Kasabov, "Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning," *IEEE Trans. Syst., Man, Cybern., Part B, Cybern.*, vol. 31, no. 6, pp. 902–918, Dec. 2001.
[36] T. Javan and Q. Chai, "A BCM theory of meta plasticity for online self reorganizing fuzzy-associative learning," *IEEE Trans. Neural Netw.*, vol. 21, no. 6, pp. 985–1003, Jun. 2010.
[37] N. Wang, M. J. Er, and X. Meng, "A fast and accurate self organizing scheme for parsimonious fuzzy neural network, *Neurocomputing*, vol. 72, nos. 16–18, pp. 3818–3829, 2009.
[38] R. J. Oentaryo, M. J. Er, L. San, L.-Y. Zhai, and X. Li, "Bayesian ART-based fuzzy inference system: A new approach to prognosis of machining process," in *Proc. IEEE Annu. Conf. Prognostic Health Soc.*, Jun. 2011, pp. 1–10.
[39] H. Han and J. Qiao, "A self-organizing fuzzy neural network based on a growing-and-pruning algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 6, pp. 1129–1143, Dec. 2010.
[40] J. D. J. Rubio, "SOFMLS: Online self-organizing fuzzy modified least squares network," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 6, pp. 1296–1309, Dec. 2009.
[41] E. Lughofer, V. Macian, C. Guardiola, and E. P. Klement, "Identifying static and dynamic prediction models for NOx emission with evolving fuzzy system," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2487–2500, 2011.
[42] C.-T. Lin, and C. S. G. Lee, *Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
[43] N. K. Kasabov and Q. Song, "DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 144–154, Apr. 2002.
[44] A. Bifet, G. Holmes, R. Kirkby, and B. P. Fahringer, "MOA: Massive online analysis," *J. Mach. Learn. Res.*, vol. 11, pp. 1601–1604, May 2010.
[45] C. F. Juang and C. T. Lin, "An on-line self-constructing neural fuzzy inference network and its applications," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 1, pp. 12–32, Feb. 1998.
[46] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
[47] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man. Cybern. Part B, Cybern.*, vol. 23, no. 3, pp. 665–684, May/Jun. 1993.
[48] M. J. Watts, "A decade of Kasabov's evolving connectionist systems: A review," *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, vol. 39, no. 3, pp. 253–269, May 2009.
[49] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, "Learn++: An incremental learning algorithm for supervised neural networks," *IEEE Trans. Syst., Man Cybern. Part C, Appl.*, vol. 31, no. 4, pp. 497–508, Nov. 2001.
[50] M. D. Muchbailer, A. Topalis, and R. Polikar, "Learn++.NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 152–168, Jan. 2009.
[51] R. Elwell and R. Polikar, "Incremental learning of concept drift in nonstationary environments," *IEEE Trans. Neural Netw.*, vol. 22, no. 11, pp. 1517–1531, Oct. 2011.
[52] C. F. Juang and Y. W. Tsao, "A self-evolving interval type-2 fuzzy neural network with on-line structure and parameter learning," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 6, pp. 1411–1424, Dec. 2008.
[53] C. F. Juang, R. B. Huang, and Y. Y. Lin, "A recurrent self-evolving interval type-2 fuzzy neural network for dynamic system processing," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 5, pp. 1092–1105, Oct. 2009.
[54] C. F. Juang, T. C. Chen, and W. Y. Cheng, "Speedup of implementing fuzzy neural networks with high-dimensional inputs through parallel processing on graphic processing units," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 717–728, Aug. 2011.
[55] G. D. Wu, Z. W. Zhu, and P. H. Huang, "A TS-type maximizing-discriminability-based recurrent fuzzy network for classification problems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 2, pp. 339–352, Apr. 2011.
[56] W. N. Street and Y. Kim, "A streaming ensemble algorithm (SEA) for large-scale classification," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, 2001, pp. 377–382.
[57] J. Z. Kolter and M. A. Maloof, "Dynamic weighted majority: An ensemble method for drifting concepts," *J. Mach. Learn. Res.*, vol. 8, pp. 2755–2790, Dec. 2007.

Mahardhika Pratama (S'12) was born in Surabaya, Indonesia. He received the B.Eng (Hons.) degree in electrical engineering from the Sepuluh Nopember Institute of Technology, Surabaya, in 2010, and the M.Sc. degree in computer control and automation from Nanyang Technological University, Singapore, in 2011. He is currently pursuing the Ph.D. degree with the University of New South Wales, South Wales, Australia.

His current research interests include machine learning, computational intelligence, evolutionary computation, fuzzy logic, neural network, and evolving adaptive systems.

Dr. Pratama has achieved the Prestigious Engineering Achievement Award from the Institute of Engineer, Singapore. He received the Best and Most Favorite Final Project. He has been nominated to *Who is Who in the World* by Marquis in 2013. He is a member of the IEEE Computational Intelligent Society and the IEEE System, Man and Cybernetic Society, and Indonesian Soft Computing Society and is an active reviewer in some top journals, such as IEEE TRANSACTIONS ON SYSTEM, MAN AND CYBERNETICS PART-B: CYBERNETICS, *Neurocomputing*, and *Applied Soft Computing*.

Sreenatha G. Anavatti received the Bachelor of Engineering degree in mechanical engineering from Mysore University, Mysore, India, in 1984, and the Ph.D. degree in aerospace engineering from the Indian Institute of Science, Bangalore, India, in 1990.

He is currently a Senior Lecturer with the School of Engineering and Information Technology, University of New South Wales, Australian Defense Force Academy, South Wales, Australia. His current research interests include control systems, flight dynamics, robotics, aeroelasticity, artificial neural networks, fuzzy systems, and unmanned systems.

Plamen P. Angelov (SM'04) is with the Computational Intelligence and Coordinator with the Intelligent Systems Research, Infolab21, Lancaster University, Lancaster, U.K. He has internationally recognized pioneering results into on-line and evolving methodologies and algorithms for knowledge extraction in the form of human-intelligible fuzzy rule-based systems and autonomous machine learning. His current research interests include the competitiveness of the industry, defense and quality of life.

He is the founding Editor-in-Chief of the Springer's journal on *Evolving Systems* and serves as an Associate Editor of several other international journals. He also chairs annual conferences organized by the IEEE as a Visiting Professor in Brazil, Germany, Spain. He is a Chair of the Technical Committee on Evolving Intelligent Systems, Systems, Man and Cybernetics Society, IEEE. He is a member of the U.K. Autonomous Systems National TC, the Autonomous Systems Study Group, NorthWest Science Council, U.K., and the Autonomous Systems Network of the Society of British Aerospace Companies. He was recognized by The Engineer Innovation and Technology Award in 2008 in two categories, Aerospace and Defense, and The Special Award.

Edwin Lughofer received the Ph.D. degree from the Department of Knowledge-Based Mathematical Systems, University of Linz, Linz, Austria.

He is currently a Post-Doctoral Fellow with the University of Linz. He has participated in several research projects on both the European and national levels. He has published more than 80 journal and conference papers in the fields of evolving fuzzy systems, machine learning and vision, clustering, fault detection and human-machine interaction, including a monograph on *Evolving Fuzzy Systems* (Springer, Heidelberg, 2011) and an edited book on *Learning in Non-Stationary Environments* (Springer, New York, 2012).

Dr. Lughofer acts as a reviewer in peer-reviewed international journals and as co-organizer of special sessions and issues at international conferences and journals. He is an Editorial Board Member and an Associate Editor of the *International Springer Journal "Evolving Systems"* and the *Elsevier Journal "Information Fusion,"* and the *European Society for Fuzzy Logic and Technology* Working Group on Learning and Data Mining. Currently, he serves as a key researcher for the national K-project "Process Analytical Chemistry" (18 partners).