An Incremental Interval Type-2 Neural Fuzzy Classifier

Mahardhika Pratama, Jie Lu, Guangquan Zhang
Centre of Quantum Computation and Intelligent System, Faculty of Engineering and Information Technology,
University of Technology Sydney, Australia,
email: pratama@ieee.org,jie.lu@uts.edu.au,guangquan.zhang@uts.edu.au

***Abstract*—Most real world classification problems involve a high degree of uncertainty, unsolved by a traditional type-1 fuzzy classifier. In this paper, a novel interval type-2 classifier, namely Evolving Type-2 Classifier (eT2Class), is proposed. The eT2Class features a flexible working principle built upon a fully sequential and local working principle. This learning notion allows eT2Class to automatically grow, adapt, prune, recall its knowledge from data streams in the single-pass learning fashion, while employing loosely coupled fuzzy sub-models. In addition, eT2Class introduces a generalized interval type-2 fuzzy neural network architecture, where a multivariate Gaussian function with uncertain non-diagonal covariance matrixes constructs the rule premise, while the rule consequent is crafted by a local non-linear Chebyshev polynomial. The efficacy of eT2Class is numerically validated by numerical studies with four data streams characterizing non-stationary behaviors, where eT2Class demonstrates the most encouraging learning performance in achieving a tradeoff between accuracy and complexity.**

Keywords—Evolving Fuzzy Systems, Fuzzy Neural Networks, Type-2 Fuzzy Systems, Sequential Learning.

I. INTRODUCTION

Zadeh [1] extended the type-1 fuzzy set to the type-2 fuzzy set, which can be envisioned as a fuzzy-fuzzy set. Unlike the type-1 fuzzy set relying on a crisp fuzzy set, the type-2 fuzzy system features a fuzzy membership degree. This strategy is proven useful to confront uncertainty in identifying an exact membership function for fuzzy system [28]. As a matter of fact, the uncertainty problem is attributed by four causes including disagreement of expert knowledge, noisy data, noisy measurement and inexact fuzzy rule parameter [2]. Nevertheless, an interval type-2 fuzzy system [3], assuming the secondary grade to be unity, is more widely studied in the literature, because all calculations are easy to perform and type-1 fuzzy mathematics are applicable. Moreover, the best representation of secondary fuzzy sets of the pure type-2 fuzzy system is still under ongoing investigation.

Evolving Fuzzy Systems (EFSs) [4], synergizing an open rule base principle and a sequential working framework, are appealing to be applied in various applications due to its underlying ability to overcome various concept drifts in the data streams. Nonetheless, most EFSs are still devised in the context of the type-1 fuzzy system [5]-[8].

The interval type-2 fuzzy concept was actualized in the context of fuzzy neural networks [9],[10], where the type reduction mechanism relies on the iterative Karnik and Mendel (KM) method. Specifically, it makes use of an interval-valued fuzzy set, which is a special case of interval type-2 fuzzy system using an interval primary membership [30]. Nevertheless, the KM method is inapplicable for the strictly sequential learning environments because of its multi-pass procedure. To remedy this bottleneck, the so-called q design factor was proposed in [11] to undertake the type reduction mechanism in lieu of the KM procedure. This solution is confirmed by the fact where the KM procedure can be substituted by the so-called minimax uncertainty bounds and the design factor q is exploited to control the minimax uncertainties. Moreover, this concept was successfully implemented in the compensatory interval type-2 fuzzy neural network in [12]. Although some seminal interval type-2 fuzzy neural networks have been proposed in [9]-[13], these works endure three major shortfalls: 1) they exploit the distance-based clustering methods to partition the input space, which is vulnerable with outliers. 2) They are not equipped by the rule base simplification procedure. 3) They are mainly targeted for solving regression problems, whereas the use of the interval type-2 fuzzy system for the classification problems still lacks investigation

This paper presents a novel evolving interval type-2 classifier, termed Evolving Type-2 Classifier (eT2Class). The eT2Class utilizes a holistic concept of evolving system, which is capable of automating the knowledge acquisition, adaptation and simplification processes in the single-pass and local learning modes. Algorithmic development of eT2Class conveys six salient learning constituents as follows: 1) eT2Class features a state-of-the art interval type-2 neural-fuzzy topology, where the rule premise is underpinned by the interval-valued multivariate Gaussian function, while the rule consequent is underpinned by a subset of the nonlinear Chebysev polynomial. 2) The knowledge building process is governed by two novel rule growing scenarios to autonomously evolve the fuzzy rules. 3) Once a new fuzzy rule is recruited, its parameters are initialized by the so-called class overlapping criterion to sidestep an overlapping fuzzy region created by a new input partition. 4) Two new rule base simplification procedures are put forward to alleviate the rule base burden. 5) eT2Class is endued by a novel rule merging technology to boost the fuzzy rule transparency, while reducing the rule base complexity. 6) Lastly, Fuzzily Weighted Generalized Least Square (FWGRLS) method is explored to adjust the rule consequents and the zero order density maximization concept [14] is used in adapting the design coefficients ql and qr. The viability of eT2Class is numerically confirmed by four data streams and comparisons with state-of-the art classifiers, where our classifier produces the most reliable classification accuracy, while retaining the lowest complexity.

The remainder of this paper is structured as follows: Section 2 details the inference scheme of eT2Class, and Section 3 elaborates algorithmic development of eT2Class. Section 4 outlines the empirical study and the last section concludes this paper.

II. ARCHITECTURE OF eT2CLASS

The rule antecedent of the generalized interval type-2 fuzzy rule is driven by the multivariate Gaussian function with uncertain inverse covariance matrix, which is capable of triggering the non-axis parallel ellipsoidal cluster.

$$\tilde{R}_i = \exp(-(X_n - C_i)\tilde{\Sigma}_i^{-1}(X_n - C_i)^T), \tilde{\Sigma}_i^{-1} = [\Sigma_{i,1}^{-1}, \Sigma_{i,2}^{2}] \quad (1)$$

where $\tilde{R}_i \in [\underline{R_i}, \overline{R_i}]$ denotes an uncertain rule firing strength. $\tilde{\Sigma}_i^{-1} = [\Sigma_{i,1}^{-1}, \Sigma_{i,2}^{-1}] \in \Re^{u \times u}$ shows an uncertain non-diagonal covariance matrix $\det(\Sigma_{i,1}) > \det(\Sigma_{i,2})$, because the upper fuzzy rule possesses a higher volume than the lower part to form a region of uncertainty. Furthermore, every element of the inverse covariance matrix pinpoints the inter-relation of input variables, which is instrumental to construct the rotation of non-axis parallel ellipsoidal cluster. Conversely, $C_i \in \Re^{1 \times u}$ denotes the centroid of *i-th* cluster and u labels the number of input dimension. Overall, the generalized fuzzy rule crafts a more reliable fuzzy region to capture the data distribution especially when streaming data are not scattered in the main axes rather contain some sorts of rotations. Because the cluster does not only span in the main axes, it can suppress the fuzzy rule demand to warrant the rule base completeness. Such merit virtually counterbalances the possible increase of rule base parameters, thus eventually arriving at comparable memory demand. Another advantage of this rule premise is its property in encapsulating the information of the input variable interaction, generally overlooked by the classical fuzzy rules.

The major drawback of the multivariate Gaussian function in the context of the interval type-2 fuzzy system is however that of its high dimensional operation, which does not yet reveal the representation of the fuzzy rule in the feature level (fuzzy set) thereby obscuring the rule semantics. To this end, the transformation of the non-axis parallel ellipsoidal cluster into its commensurate fuzzy set is undertaken by applying the second method of [18] as follows:

$$\tilde{\sigma}_i = \frac{r}{\sqrt{\tilde{\Sigma}_{ii}}}, \tilde{\sigma}_i \in [\sigma_i^1, \sigma_i^2] \quad (2)$$

where $\tilde{\sigma}_i \in \Re^{1 \times u}$ stands for the uncertain radii of the *i-th* rule extracted from the non-diagonal covariance matrix. Note that this method features a swift operation to elicit the fuzzy set radii. However, this approach is slightly inaccurate when the cluster revolves around 45 degree. On the contrary, the centroid of the fuzzy set is akin to that of the non-axis parallel ellipsoidal cluster, in which it allows to be fed directly to the inference scheme. Once obtaining the upper and lower radii (2), the membership degree of the interval type-2 fuzzy set with uncertain Standard Deviations (SDs) [3] can be enumerated as follows:

$$\tilde{\mu}_{i,j} = \exp(-(\frac{x_j - c_{i,j}}{\tilde{\sigma}_{i.j}})^2), \ \tilde{\sigma}_{i,j} \in [\sigma^1{}_{i,j}, \sigma^2{}_{i,j}] \quad (3)$$

Henceforth, the upper and lower spatial firing strength of *i-th* rule $\tilde{R}_i = [\underline{R_i}, \overline{R_i}]$ can be produced with the help of the *t-norm* operator as follows:

$$\underline{R_i} = \prod_{j=1}^{u} \underline{\mu_{i,j}}, \overline{R_i} = \prod_{j=1}^{u} \overline{\mu_{i,j}} \quad (4)$$

The major deficiency of the standard Takagi Sugeno Kang (TSK) rule consequent is that it does not fully exploit a local approximation aptitude. To correct this shortcoming, the rule consequent of the eT2Class fuzzy rule is built upon a subset of the Chebyshev polynomial [19], expanding the degree of freedom of the rule consequent to rectify the local mapping characteristic. Generally speaking, the output node of eT2Class is formed as follows:

$$y_i^o = x_e \tilde{\Omega}_i, \tilde{\Omega}_i \in [\Omega_i^l, \Omega_i^r] \quad (5)$$

where y_i^o stands for the regression output of the *i-th* rule falling into the *o-th* class, while $\tilde{\Omega}_i \in \Re^{(2u+1) \times m}$ denotes the weight vector and m stands for the number of output dimension. In what follows, the weight vector comprises L and R components $\tilde{\Omega}_i \in [\Omega_i^l, \Omega_i^r]$, where $\Omega_i^l = [w_1^{l,i,o}, ..., w_{2u+1}^{l,i,o}]$ and $\Omega_i^r = [w_1^{r,i,o}, ..., w_{2u+1}^{r,i,o}]$. By means of this rule consequent, a sort of interval uncertainty is instilled in the rule consequent. Furthermore, $x_e \in \Re^{1 \times (2u+1)}$ indicates an extended input vector, which is yielded by a non-linear mapping based on up to second order of the Chebyshev polynomial as follows:

$$T_{n+1}(x_j) = 2x_j T_n(x_j) - T_{n-1}(x_j) \quad (6)$$

where $T_0(x_j) = 1, T_1(x_j), T_2(x_j) = 2x^2{}_j - 1$. For example, if we have 2-D input dimension $X = [x_1, x_2]$, the expanded input vector is formed as $x_e = [1, T_1(x_1), T_2(x_1), T_1(x_2), T_2(x_2)]$. Other functional link rule outputs like trigonometric or polynomial function also work perfectly with eT2Class. Nevertheless, they impose more considerable memory demand. The term 1 is included in the expanded input vector to include the intercept of the rule consequent. Otherwise, the rule consequent may land on origin, which leads to untypical gradient.

The design coefficients $[q_l, q_r]$ of [13] are employed to perform the type reduction mechanism, where the crux is to govern the proportion of upper and lower outputs $[y_l, y_r]$. The lower and upper outputs are formulated as follows:

$$y_{l,o} = \frac{(1-q_l^o)\sum_{i=1}^{P} \underline{R_i} y_{i,o}^l + q_l^o \sum_{i=1}^{P} \overline{R_i} y_{i,o}^l}{\sum_{i=1}^{P} \underline{R_i} + \overline{R_i}} \quad (7)$$

$$y_{r,o} = \frac{(1-q_r^o)\sum_{i=1}^{P} \overline{R_i} y_{i,o}^r + q_r^o \sum_{i=1}^{P} \underline{R_i} y_{i,o}^r}{\sum_{i=1}^{P} \underline{R_i} + \overline{R_i}} \quad (8)$$

where P denotes the number of fuzzy rules. In essence, the lower and upper outputs $[y_l, y_r]$ portray type reduced sets, which produce the crisp output as $y_o = y_{l,o} + y_{r,o}$. If the decision boundary is assembled by the MIMO architecture, eT2Class outputs the classification decision via the maximum operator as follows:

$$y = \max_{o=1,..,m}(y_o) \quad (9)$$

It is worth-mentioning that other classifier architectures are also compatible with eT2Class. The MIMO architecture is put forward herein, because it is prominent in the literature, thus being able to support fair comparisons with existing classifiers.

III. LEARNING POLICY OF ET2CLASS

1) Rule Generation Mechanism: two rule growing scenarios are devised to delve the generalization potential and the summarization power of streaming data. The first rule growing cursor has its root in [6], which is tailored to accommodate the interval type 2 fuzzy system. The centric construct of this method is to foresee the statistical contribution of streaming data, which in pinciple indicates the future contribution of streaming data. Assuming that the training samples are unifornly distributed, we can prognosticate the statistical contribution of a hypothetical cluster (P+1[st]) as follows:

$$DS_n = \frac{1}{2}\left(\frac{V_{P+1}^{1}}{\sum_{i=1}^{P+1} V_i^1} + \frac{V_{P+1}^{2}}{\sum_{i=1}^{P+1} V_i^2}\right) \quad (10)$$

where V_i designates the volume of *i-th* rule. In short, the hypothetical rule is desired to be a new rule given that it conveys a higher volume than the existing ones as follows:

$$(V_{P+1}^{1} + V_{P+1}^{2}) > \max_{i=1,..,P}(V_i^1 + V_i^2) \quad (11)$$

Another method is amalgamated to monitor the contribution of the streaming data, which customizes that of [6] to serve the interval type-2 fuzzy system. The key idea of this method is to quantify the spatial proximity of an incoming datum against all preceding training patterns without storing all data in the memory. This facet is equivalent with a data density measure. Because we deal with the interval type-2 fuzzy sets with uncertain SDs, owning an identical centroid, this method is defined as follows:

$$DQ_N = \sqrt{\frac{U_n}{U_n(1+b_n) - 2h_n + g_n}} \quad (12)$$

$U_n = U_{n-1} + DQ_{N-1}$, $b_n = \sum_{j=1}^{u}(x_j^N)^2$, $h_n = \sum_{j=1}^{u} x_j^N p_n^j$,

$p_n = p_{n-1} + DQ_{N-1}X_n$ $g_n = g_{n-1} + DQ_{N-1}b_n$

where all recursive parameters are initialized as zero prior to the training process. A datum is deemed appealing to serve as a new rule, when it occupies a dense fuzzy region or a remote area from the influence zone of the existing fuzzy rules as follows:

$$DQ_N \geq \max_{i=1,..,P}(DQ_i) \text{ or } DQ_N \leq \min_{i=1,...,P}(DQ_i) \quad (13)$$

Albeit offering the information of the data density with a trivial computational complexity, this method is anticipated to suffer from an outlier's stumbling block. To this end, we engage the weighting factor DQ_{N-1} herein to diminish the outlier's leverage [15]. One can envision that the $DQ_N \leq \min_{i=1,...,P}(DQ_i)$ criterion should be equipped by the rule pruning mechanism to forestall outliers to be engaged in the rule base. The major rationale is that outliers are normally situated in the remote area, thus possibly being endorsed by this criterion to act as a fuzzy rule. On the contrary, the $DQ_N \geq \max_{i=1,..,P}(DQ_i)$ condition ought to be circumspectly cultivated, because a newly created fuzzy region presumably induces an overlapping region owing to an imminent relationship with inter-class clusters.

2) Initializatin of Fuzzy Rule: In [16], the distance ratio between intra-and intra clusters are made use of measuring the risk of the class overlapping situation. The so-called potential per-class method is designed to circumvent the class overlapping circumstance in [7]. Nevertheless, the technical flaw of the distance ratio method is that it does not account an unpurified cluster issue. By extension, the potential per-class method is under-explored for the interval type-2 fuzzy system, so that it is worth-studying in this paper.

Intrinsically, the class overlapping situation most likely ensues, when the recruitment of a new fuzzy rule is ascertained by $DQ_N \geq \max_{i=1,..,P}(DQ_i)$. In what follows, a new fuzzy rule is populated by most other samples. If $DQ_N \geq \max_{i=1,..,P}(DQ_i)$ is observed, we activate the potential per-class method as follows:

$$DQ_o = \sqrt{\frac{(N_o - 1)}{(N_o - 1)(ab_n + 1) + cb_{no} - 2bb_{no}}} \quad (14)$$

$ab_n = \sum_{j=1}^{u+m}(x_j^N)^2$, $cb_{no} = cb_{no-1} + \sum_{j=1}^{u+m}(x_j^{no-1})^2$, $bb_{no} = \sum_{j=1}^{u+m} x_j^N d_j^{no}$,

$d_{no} = d_{no-1} + x_{no-1}$

where x_j^N stands for the latest incoming datum and x_j^{no} denotes the latest datum falling into the *o-th* class, whereas N_o labels the number of samples falling into the *o-th* class. The most vulnerable condition, which may slacken the classifier's generalization, is given by $\max_{o=1,..,m}(DQ_o) \neq true\_class\_label$. Note that this condition possibly aggravates the non-linearity degree of the decision boundary, since a substantial adjustment of the decision boundary should be made to cope with the class overlapping case. A plausible avenue to cope with this situation is to shrink a coverage span of a new cluster and is to shift a cluster's focal-point away from the existing data clouds. Suppose that *ie* is the nearest inter-class cluster and *ir* is the nearest intra-class cluster, the new rule is crafted as follows:

$$c_{P+1,j}^{1,2} = x_j - \rho_2 (c_{ie,j}^{1,2} - x_j) \;\; dist_1^j = \frac{\rho_1}{N_{win}} \left| c_{P+1,j}^{1} - c_{ie,j}^{1} \right|$$

$$\tilde{\Sigma}_{p+1}^{-1} = (dist_{1,2}^T dist_{1,2})^{-1} \;,\; dist_2^j = \frac{2\rho_1}{N_{win}} \left| c_{P+1,j}^{2} - c_{ie,j}^{2} \right| \quad (15)$$

where $\rho_1 = r_{ir}^j / r_{ie}^j$ stands for an overlapping factor, steering the size of the fuzzy region. It is fixed as a distance ratio between the winning intra-class cluster and the winning inter-class cluster. Obviously, this setting is plausible in respect to the fact, where the cluster's coverage span should be diminished as the similarity degree of the winning inter-class cluster. Conversely, $\rho_2 \in [0.01 - 0.1]$ labels a shifting factor, which is stipulated as 0.01 in this paper. Because we deal with uncertain non-diagonal covariance matrixes, we integrate $N_{win}, N_{win}/2$ in tailoring the footprint of uncertainty.

Another condition shown by $\max_{o=1,..,m} (DQ_o) = true\_class\_label$ is easier to manage than the former one, because a newly crafted rule has more neighboring relationship to the intra-class data clouds than that of inter-class ones. Accordingly, it does not affect the decision surface, so that a negligible chance of misclassification is incurred. We can then assign more confident parameters to a new rule as follows:

$$c_{P+1,j}^{1,2} = x_j + \rho_2 (c_{ir,j}^{1,2} - c_{ie,j}^{1,2}) \;,\; dist_1^j = \frac{\rho_1}{N_{win}} \left| c_{P+1,j}^{1} - c_{ir,j}^{1} \right|$$

$$dist_2^j = \frac{2\rho_1}{N_{win}} \left| c_{P+1,j}^{2} - c_{ir,j}^{2} \right| \;,\; \tilde{\Sigma}_{p+1}^{-1} = (dist_{1,2}^T dist_{1,2})^{-1} \quad (16)$$

It is worth noting that even though the new fuzzy rule can end up with a significantly overlapping position with the intra-class cluster in the future, this situation can be overcome by undertaking the rule merging scenario, deliberated elsewhere in this paper.

If the new fuzzy rule portrays an uncharted territory from the existing fuzzy rules $DQ_N \leq \min_{i=1,...,P} (DQ_i)$, the new fuzzy rule is evolved as follows:

$$c_{P+1,j}^{1,2} = x_j \;,\; dist_1^j = \frac{\rho_1}{N_{win}} \left| x_j - c_{ir,j}^{1} \right| ,$$

$$dist_2^j = \frac{2\rho_1}{N_{win}} \left| x_j - c_{ir,j}^{2} \right| \;,\; \tilde{\Sigma}_{p+1}^{-1} = (dist_{1,2}^T dist_{1,2})^{-1} \quad (17)$$

Noticeably, the fuzzy rule initialization merely evokes a subtle impact to the evolution of other clusters, because it snapshots a remote fuzzy region.

The rule consequent of the new fuzzy rule is likewise constructed as the winning rule as follows:

$$\Omega_{P+1}^{l,r} = \Omega_{win}^{l,r} \;,\; \Psi_{P+1}^{l,r} = \omega I \quad (18)$$

where ω denotes a large positive constant, fixed as $\omega = 10^5$. It is borne out in [17] that a real solution as produced by the batched learning scheme can be incrementally emulated provided that Ψ_{P+1} is determined as a big positive definite matrix. On the other hand, the underlying rationale of the initialization of the rule consequent is that the winning rule is supposed to portray the pertinent data trend with the new rule.

3) Winning rule selection method: In eT2Class, the winning rule is extracted in Bayesian sense, where a cluster, possessing the maximum posterior probability $win = \arg\max_{i=1,...,P} \hat{P}(R_i|X)$, is opted as the winning cluster. The peculiar property of this method is that the prior probability of the clusters is impactful in chosing the winning rule. This technique conveys a probabilistic standpoint to select the most compatible rule, when there are some clusters sitting on par to the datum. For brevity, the posterior, prior probabilities and likelihood function are respectively expressed as follows:

$$\hat{P}(R_i|X) = \frac{1}{2}\left(\frac{\hat{p}(X|R_i)^1 \hat{P}(R_i)}{\sum_{i=1}^{P} \hat{p}(X|R_i)^1 \hat{P}(R_i)} + \frac{\hat{p}(X|R_i)^2 \hat{P}(R_i)}{\sum_{i=1}^{P} \hat{p}(X|R_i)^2 \hat{P}(R_i)}\right) \quad (19)$$

$$\hat{P}(X|R_i)^{1,2} = \frac{1}{(2\pi)^{1/2} V_{i,1,2}^{1/2}} \exp(-(X - C_i^{1,2}) \Sigma_{i,1,2}^{-1} (X - C_i^{1,2})^T) \quad (20)$$

$$\hat{P}(R_i) = \frac{\log(N_i + 1)}{\sum_{i=1}^{P} \log(N_i + 1)} \quad (21)$$

where N_i exhibits the population of the *i-th* cluster. One can envisage that the prior probability $\hat{P}(R_i)$ is softened from its original definition to allow the newly added fuzzy rules to be competitive in the category choice phase.

4) Rule premise adaptation: the rule premise of the winning cluster is fine-tuned, if the rule growing criteria in (11),(13) are not complied. In what follows, the new datum solely incurs a marginal conflict level, thus merely soliciting the rule premise adaptation to refine the coverage span of the winning cluster. Specifically, the rule premise of the winning rule is adjusted as follows:

$$C_{win,1,2}^{N} = \frac{N_{win}^{N-1}}{N_{win}^{N-1} + 1} C_{win,1,2}^{N-1} + \frac{(X_N - C_{win,1,2}^{N-1})}{N_{win}^{N-1} + 1} \quad (23)$$

$$\Sigma_{win,1,2}(N)^{-1} = \frac{\Sigma_{win,1,2}(N-1)^{-1}}{1-\alpha} + \frac{\alpha}{1-\alpha} \quad (24)$$
$$\frac{(\Sigma_{win,1,2}(N-1)^{-1}(X_N - C_{win,1,2}^{N-1}))(\Sigma_{win,1,2}(N-1)^{-1}(X_N - C_{win,1,2}^{N-1}))^T}{1 + \alpha(X_N - C_{win,1,2}^{N-1})\Sigma_{win,1,2}(old)^{-1}(X_N - C_{win,1,2}^{N-1})^T}$$

$$N_{win}^{N} = N_{win}^{N-1} + 1 \quad (25)$$

where $\alpha = 1/(N_{win}^{N-1} + 1)$. In principle, the rule premise adaptation scheme emanates from the sequential maximum likelihood estimation for the spherical cluster amended to gear the non-axis parallel ellipsoidal cluster requirement. Furthermore, (24) is revamped to enable a direct update of the inverse covariance matrix to sidestep a re-inversion step, which can cause numerical stability due to an ill-defined matrix and can then retard the training process.

5) Rule Pruning Technology: eT2Class is endued by two rule pruning scenarios which are capable of detecting both obsolete and inconsequential clusters. The inconsequential rules are those playing little role during their lifespan. This situation means outliers misleadingly recruited as fuzzy rules in earlier training observations. The first method is in charge of capturing such rules, and is written as follows:

$$ERS_i^{\,n}=\frac{1}{2}\left|\sum_{j=1}^{2u+1}\sum_{o=1}^{m}\Omega_{i,l}^{\;o,j}+\Omega_{i,r}^{\;o,j}\right|\left(\frac{V_i^1}{\sum_{i=1}^{p}V_i^1}+\frac{V_i^2}{\sum_{i=1}^{p}V_i^2}\right)\quad(26)$$

The crux of the method is to oversee the statistical contribution of the fuzzy rules, signified by their volumes. Apart from that, the first method also fathoms the contribution of the rule consequent, because a minor output weight influences marginally to the system output. In a nutshell, the fuzzy rules are pruned, when satisfying $ERS_i^{\,n}\le mean(ERS_i^{\,n})-std(ERS_i^{\,n})$.

In addition, the fuzzy rules should be removed from the rule base given that they are outdated due to concept drifts, because such fuzzy rule is no longer relevant to cover the data trend. To this end, the rule pruning method should be able to trace the cluster evolution in respect to the data distribution as follows:

$$\chi_i=\sqrt{\frac{(N-1)\chi_{N-1,i}^{\;\;2}}{2\chi_{N-1,i}^{\;\;2}+2\chi_{N-1,i}^{\;\;2}\sum_{j=1}^{m+o}(x_{i,j}^{\;N-1}-c_{i,j})^2+(N-2)}}\quad(27)$$

Note that (27) is derived under the assumption of the identical upper and lower centroids. Nonetheless, (27) can suit to the case of uncertain means easily. In short, the fuzzy rules are trimmed from the rule base, should $\chi_i^{\,n}\le mean(\chi_i^{\,n})-std(\chi_i^{\,n})$ be satisfied. The T2P+ and T2ERS methods are extended from its type-1 version in [6]. Because the obsolete rules are deactivated by the second method, it is required to enable the rule recall mechanism to surmount the cyclic concept drift, which is prevalent in real-world data streams. That is, the recurring concept drift reveals a situation, in which the past data distribution re-emerges again in the future. Consequently, the cyclic concept drift substantiates the validity of previously pruned fuzzy rules to be appended. Note that adding completely new fuzzy rule to manage the cyclic concept drift without paving likelihood to reignite the old fuzzy rules undermines the logic of evolving system. The pruned fuzzy rules are revived, if $\max_{i^*=1,..,P^*}(\chi_{i^*})>\max_{i=1,..,P+1}(DQ_i)$ is observed, where P^* stands for the number of fuzzy rules eliminated by the T2P+ method thus far. In this context, the new fuzzy rule is assigned as that of a pruned fuzzy rule as follows:

$$C_{P+1,1,2}=C_{i^*,1,2},\Sigma_{P+1,1,2}^{\;-1}=\tilde{\Sigma}_{i^*,1,2}^{\;-1},\tilde{\Psi}_{P+1}^{\;l,r}=\tilde{\Psi}_{i^*}^{\;l,r},\Omega_{P+1}^{\;l,r}=\Omega_{i^*}^{\;l,r}\quad(28)$$

In essence, the fuzzy rules, gotten rid of the T2P+ method, are confined to solely cultivate (27) without being engaged in any learning scenarios. Accordingly, the computational complexity is still alleviated.

6) Rule Merging Scenario: the rule merging scenario plays a crucial role to guaranteee the compactness and interpretability of the rule base. In realm of IT2FNNs [18],[19], the rule merging scenario has been incorporated by means of the distance-based or shape-based similarity measure. Nevertheless, these pioneering works do not take into account both distance and shape altogether in abstracting the similarity of two clusters. To correct the deficiency of its predecessors, the rule merging scenario of eT2Class is built upon the seminal work of Wu and Mendel in [20] termed the vector similarity method. The salient feature of this approach is that the similarity measure of two clusters is enumerated in respect to their distance and countour, thus allowing to quantify the similarity degree more accurately. In addition, the blow-up effect as a result of merging non-homogeneous clusters is another important issue to be taken into account, because merging two different orientation clusters may result in an inexact representation of the merged cluster. In short, an over-sized cluster, which is prone to trigger the cluster delamination situation, is imposed by the rule merging mechanism. Note that the cluster delamination situation, which harms the classifier's accuracy portrays an over-sized cluster covering two or more distinguishable data clouds. To this end, the rule merging strategy of eT2Class is endued by the blow up check, monitoring the volume of the cluster to stave off two non-homogeneous clusters to be coalesced.

The crux of the vector similarity method is to vet the similarity of two clusters based on their proximities and their shapes in one joint formula as follows:

$$s_{v,j}(win,i)=s_{1,j}(win,i)\times s_{2,j}(win,i)\quad(29)$$

where $s_{1,j}(win,i)\in[0,1]$ denotes the shape-based similarity measure, while $s_{2,j}(win,i)\in[0,1]$ stands for the distance-based similarity measure. The peculiar characteristic of this method lies on its alignment procedure as a consequence of accounting the shape and distance-based similarity measures separately. In what follows, the focal point of two clusters is aligned, such that they coincide $c_{win,j}^{\;1,2}=c_{i,j}^{\;1,2}$. Furthermore, the Jaccard similarity measure is extended by amalgamating the average cardinality concept to accommodate the interval type-2 fuzzy rule as follows:

$$s_{1,j}(win,i)=\frac{M(\underline{\mu_{win,j}}\cap\underline{\mu_{i,j}})+M(\overline{\mu_{win,j}}\cap\overline{\mu_{i,j}})}{M(\underline{\mu_{win,j}}\cup\underline{\mu_{i,j}})+M(\overline{\mu_{win,j}}\cup\overline{\mu_{i,j}})}\quad(30)$$

where $\cap$ and $\cup$ indicate the union and intersection of two fuzzy sets $\tilde{\mu}_{win},\tilde{\mu}_i$. For simplicity, the union of two fuzzy sets can be formed as $M(\underline{\mu_{win,j}}\cup\underline{\mu_{i,j}})=M(\underline{\mu_{win,j}})+M(\underline{\mu_{i,j}})-M(\underline{\mu_{win,j}}\cap\underline{\mu_{i,j}})$. Because the Gaussian function signifies a highly nonlinear contour, the size of the fuzzy set is estimated by the triangular approach for simplicity [21] as follows:

$$M(\underline{\mu_{win,j}})=\int_{-\infty}^{\infty}\exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)dx=\sigma_{win,j}^{\;1}\sqrt{2\pi}\quad(31)$$

Therefore, the shape-based similarity measure can be established as follows:

$$M(\underline{\mu_{win,j}}\cap\underline{\mu_{i,j}})=\frac{h^2}{2}+\frac{h^2((\sigma_{win,j}^{\;1}-\sigma_{i,j}^{\;1}))}{2(\sigma_{i,j}^{\;1}-\sigma_{win,j}^{\;1})}-\frac{h^2(\sigma_{win,j}^{\;1}+\sigma_{i,j}^{\;1})}{2(\sigma_{win,j}^{\;1}-\sigma_{i,j}^{\;1})}\quad(32)$$

where $h = \max[0, x]$. We can arrive at (30) by performing the similar mathematical operations for the upper fuzzy rule $M(\overline{\mu_{win,j}} \cap \overline{\mu_{i,j}})$, $M(\overline{\mu_{win,j}} \cup \overline{\mu_{i,j}})$.

On the other hand, the distance-based similarity measure is implemented by the so-called kernel-based metric method [22], where the average cardinality approach is utilized in probing the interval type-2 fuzzy set. In a nutshell, it is defined as follows:

$$S_{2,j}(win,i) = \frac{\exp(-A)+\exp(-B)}{2} \tag{33}$$

$$A = \left|c_{win,j}^{1} - c_{i,j}^{1}\right| - \left|\sigma_{win,j}^{1} - \sigma_{i,j}^{1}\right| \quad B = \left|c_{win,j}^{2} - c_{i,j}^{2}\right| - \left|\sigma_{win,j}^{2} - \sigma_{i,j}^{2}\right|$$

where the efficacy of this method relies on the following appealing properties.

$$S_{2,j}(A,B) = 1 \Leftrightarrow |C_A - C_B| + |\sigma_A - \sigma_B| = 0 \Leftrightarrow C_A = C_B \wedge \sigma_B = \sigma_B$$

$$S_{2,j}(A,B) < \varepsilon \Leftrightarrow |C_A - C_B| > \delta \vee |\sigma_A - \sigma_B| > \delta$$

Once $s_{1,j}(win,i) \in [0,1]$ and $s_{2,j}(win,i) \in [0,1]$ are elicited, it allows to land on (29), which finalizes the vector similarity measure. For brevity, two fuzzy rules are deemed identical given that (29) surpasses a predefined threshold where the rule similarity can be produced by the *t-norm* operator as follows:

$$S_v \geq \rho_3, S_v = \min_{j=1,..,u}(s_{v,j}) \tag{34}$$

where $\rho_3 \in [0,1]$ labels a non-problem-specific threshold.

To the light of the blow up effect, the volume of two clusters to be fused is examined to capture non-homogeneous clusters where we specifically compare the volume of two independent clusters against the merged cluster. The rule merging process is withheld on condition that the volume of merged cluster beats that of two standalone clusters. This situation is defined as follows:

$$V_{merged}^{1} + V_{merged}^{2} \leq u((V_{win}^{1} + V_i^{1}) + (V_{win}^{2} + V_i^{2}) \tag{35}$$

The term u is consolidated herein to hedge the curse of dimensionality problem. By extension, the rule merging process is carried out by benefiting from the weighted average strategy as follows:

$$C_{merged}^{1,2} = \frac{C_{win}^{1,2} N_{win}^{old} + C_i^{1,2} N_i^{old}}{N_{win}^{old} + N_i^{old}} \tag{35}$$

$$\Sigma^{-1}{}_{merged}^{1,2} = \frac{\Sigma^{-1}{}_{win}^{1,2} N_{win}^{old} + \Sigma^{-1}{}_i^{1,2} N_i^{old}}{N_{win}^{old} + N_i^{old}} \tag{36}$$

$$N_{merged}^{new} = N_{win}^{old} + N_i^{old} \tag{37}$$

The unique property of such rule merging concept can be seen in which each rule is weighted by its supports. Intrinsically, the more supported rule should be more impactful to the eventual shape and orientation of the merged cluster, because the cluster's populations undoubtedly beckon the underlying data distribution. Otherwise, a loss of the cluster's supports is caused.

One can envisage the contradictory problem, where the rule antecedent of two fuzzy rules is identical, whereas the rule output of these rules is noticeably distinct. Accordingly, this leads to assemble a contradictory measure, which regulates the rule consequent merging mechanism. To the light of this issue, the similarity of the rule consequent is appraised by vetting the angle created by the two rule consequents. Because we deal with the non-linear Chebyshev function, the orientation of the nonlinear function is nothing else but delineated by that of its first order component and its higher order constituents merely fashion its non-linear oscillations. Therefore, the similarity of the rule consequent part is formulated as follows:

$$\hat{\phi}_{l,r} = \max_{o=1,...,m}(\phi_{o,l,r}) \;\; \phi_{o,l,r} = \arccos\left(\left|\frac{a_{o,l,r}^{T} b_{o,l,r}}{\|a_{o,l,r}\|\|b_{o,l,r}\|}\right|\right) \tag{38}$$

$$S_{out}(\Omega_i^{l,r}, \Omega_{i+1}^{l,r}) = \begin{cases} 1 - \frac{2}{\pi}\hat{\phi}_{l,r}, \hat{\phi}_{l,r} \in \left[0, \frac{\pi}{2}\right] \\ \frac{2}{\pi}(\hat{\phi}_{l,r} - \frac{\pi}{2}), \hat{\phi}_{l,r} \in \left[\frac{\pi}{2}, \pi\right] \end{cases} \tag{39}$$

$$a = [w^{o,l,r}{}_{win,1}, w^{o,l,r}{}_{win,3}, ..., w^{o,l,r}{}_{win,2u-1}],$$

$$b = [w^{o,l,r}{}_{i,1}, w^{o,l,r}{}_{i,3}, ..., w^{o,l,r}{}_{i2u-1}] \tag{40}$$

where $a, b \in \Re^{m \times u}$. The maximum operator is employed to figure out the maximum similarity of the rule output among classes, because it can confirm the rule output merging procedure, when the maximum angle created by the rule consequent of a particular class is high. Furthermore, the merging process is inspired by the Yager's participatory learning concept as follows:

$$\Omega_{merged}^{l,r} = \Omega_{win}^{l,r} + \gamma\delta(\Omega_{win}^{l,r} - \Omega_i^{l,r}), \gamma = \frac{N_{win}^{old}}{N_{win}^{old} + N_i^{old}} \tag{40}$$

$$\delta = \begin{cases} 1, S_v \leq S_{out} \\ 0, S_v > S_{out} \end{cases} \tag{41}$$

In conjunction with the participatory learning [23] $\gamma \in [0,1]$ can be regarded as the basic learning rate and δ stands for the compatibility measure. The arousal index is set as 0 and $N_{win} > N_i$. In essence, the rule consequents are fused given that their similarity exceeds that of the rule premise. The rule merging procedure solely takes place in the winning rule, because it is the only one granted by the rule resonance (23)-(25). Inevitably, the rule premise adaptation is the major rationale of the rule redundancy. Notwithstanding scattered adequately apart from other clusters, the cluster can be still jeopardized by the redundancy issue notably when the next training samples fill up the gap between the clusters.

7) Fuzzily Weighted Generalized Recursive Least Square (FWGRLS) Method: the FWGRLS method constitutes a local learning version of the Generalized Recursive Least Square (GRLS) method [24]. The salient trait of this method can be seen in its weight decay term, which is capable of maintaining the weight vector to merely hover arround a small bounded interval. As a result, the classifier's generalization can be intensified, while fostering the compactness of the rule base. Although it is well-established for the type-1 fuzzy system [6], the use of the FWGRLS method for the interval type-2 fuzzy system is uncharted. On the other hand, the FWGRLS method is built upon the local learning scheme, which offers a greater flexibility than that of the glocal learning procedure as omnipresent in various interval type-2 fuzzy systems. Consequently, a learning scenario of a specific fuzzy rule does not harm the convergence and stability of other fuzzy rules. In a nutshell, the FWGRLS method is expressed as follows:

$$\psi^{l,r}(n)=\Psi_i^{l,r}(n-1)F(n)\left(\frac{\Delta(n)}{\widetilde{\Lambda}_i(n)}+F(n)\Psi_i^{l,r}(n-1)F^T(n)\right)^{-1} \quad (42)$$

$$\Psi_i^{l,r}(n)=\Psi_i^{l,r}(n-1)-\psi^{l,r}(n)F(n)\Psi_i^{l,r}(n-1) \quad (43)$$

$$\Omega_i^{l,r}(n)=\Omega_i^{l,r}(n-1)-\varpi\Psi_i^{l,r}(n)\nabla\xi(\Omega_i^{l,r}(n-1))+\Psi_i^{l,r}(n)(t(n)-y(n)) \quad (44)$$

$$y(n)=x_{en}\Omega_i^{l,r}(n) \text{ and } F(n)=\frac{\partial y(n)}{\partial\Omega_i^{l,r}(n)}=x_{en} \quad (45)$$

where $\widetilde{\Lambda}_i(n)\in\Re^{(P+1)\times(P\times1)}$ denotes a diagonal matrix, whose diagonal elements comprise the firing strength of the fuzzy rule $\widetilde{R}_i$. $\Delta(n)$ stands for the output covariance matrix, which can be set as an identify matrix [24]. Moreover, $\nabla\xi(\Omega_i^{l,r}(n-1))$ shows the gradient of the weight decay function and $\varpi\approx10^{-15}$ exhibits a case-insensitive predefined constant. The gradient of the weight decay function can be stipulated as any nonlinear function, which might have an inexact gradient solution. Therefore, it is extended to *n-1* step whenever the gradient solution is over-complex to be obtained. The quadratic weight decay function is utilized by eT2Class, because it is capable of shrinking the weight vector proportionally its current values.

8) *Adaptation of The Design Factors:* eT2Class relies on the design factors $[q_l,q_r]$, steering the proportion of the upper and lower outputs $[y_l,y_r]$. The design coefficients are adapted by the so-called Zero Error Density Maximization (ZEDM) principle [14], which is confirmed to be more robust than the Mean Square Error (MSE)-based adaptation scheme. Since the ZEDM method aims to minimize the error entropy, the distance between the probability distribution of the target class and the classifier's output is dampened. Accordingly, this perspective results in the system error to culminate at origin. In practice, it is troublesome to determine the data distribution of the errror entropy. We thus exploit the parzen window estimation method to form the cost function as follows:

$$\hat{f}(0)=\frac{1}{Nh\sqrt{2\pi}}\sum_{n=1}^{N}\exp\left(-\frac{e_{n,o}^2}{2h^2}\right)=\frac{1}{Nh\sqrt{2\pi}}\sum_{n=1}^{N}K\left(\frac{-e_{n,o}^2}{2h^2}\right) \quad (46)$$

where *N* labels the number of data streams encountered thus far and *h* exhibits the smoothing factor, which is assigned as 1. $e_{n,o}$ exhibits the system error of the *n-th* training cycle of the *o-th* class. Furthermore, the gradient ascent method is deployed as follows:

$$q_{i,r}^o(N)=q_{l,r}^o(N-1)+\eta_o\frac{\partial\hat{f}(0)}{\partial q_{l,r}^o}=q_{l,r}^o(N-1)-\eta_o\frac{1}{N\sqrt{2\pi}}\sum_{n=1}^{N}K\left(-\frac{e_{n,o}^2}{2}\right)\frac{\partial E}{\partial q_{l,r}^o} \quad (47)$$

where η_o denotes the adaptive learning rate, which is allocated by virtue of the Lyapunov criterion to warrant the convergence. (47) is revamped further to satisfy the strictly sequential learning environment as follows:

$$\sum_{n=1}^{N}\exp\left(\frac{-e_n^2}{2}\right)=A_N=A_{N-1}+\exp\left(\frac{-e_{N,o}^2}{2}\right),\frac{\partial\hat{f}(0)}{\partial q_{l,r}^o}=\frac{A_N}{N\sqrt{2\pi}}\frac{\partial E}{\partial q_{l,r}^o} \quad (49)$$

The gradient term is obtained by using the chain rule as follows:

$$\frac{\partial E}{\partial q_l}=e_{N,o}\frac{\sum_{i=1}^{P}x_e\Omega_{i,o}^l(\underline{R_i}-\overline{R_i})}{\sum_{i=1}^{P}(\underline{R_i}-\overline{R_i})},\ \frac{\partial E}{\partial q_r}=e_{N,o}\frac{\sum_{i=1}^{P}x_e\Omega_{i,o}^r(-\underline{R_i}+\overline{R_i})}{\sum_{i=1}^{P}(-\underline{R_i}+\overline{R_i})}$$

On the other hand, the learning rate plays a crucial role to expedite the training process and is accordingly governed as follows:

$$\eta_o(N)=\begin{cases}\rho_5\eta_o(N-1), & \hat{f}(0)^N\geq\hat{f}(0)^{N-1}\\ \rho_4\eta_o(N-1), & \hat{f}(0)^N<\hat{f}(0)^{N-1}\end{cases},\ 0<\rho_4<1<\rho_5 \quad (48)$$

where $\rho_5\in(1,1.5]$, $\rho_4\in[0.5,1)$ stand for the learning rate factors, which govern the growth and decline of the learning rates. For brevity, they can be set as $\rho_4=1.1$, $\rho_5=0.9$ without compromising the learning performance. To assure the convergence, we foster the learning rate in the range of $0<\eta_o<\frac{2N\sqrt{2\pi}}{(P_{o,\max})^2A_N}$, which is derived from the Lyapunov stability criterion. The mathematical derivation is however omitted in this paper to comply the maximum page limit.

Table 1. consolidated results of benchmarked system in four datasets

algorithm		Checker Board	Car+	Gaussian	Hyper plane
eT2Class	CR	0.65±0.07	**0.78±0.13**	**0.74±0.0**	**0.9±0.13**
	Rule	**1.4±0.5**	**1.5±0.5**	**1.6±0.5**	**2.1±0.6**
	Time	1.9±0.5	0.17±0.06	1.6±0.2	0.4±0.07
	Rule base	**27.6±9.7**	**156±54.8**	**41±12.1**	**74±23**
pClass	CR	0.65±0.1	0.77±0.15	0.73±0.4	0.89±0.5
	Rule	3.1±1.4	2.5±0.9	3.3±1.1	6.2±4
	Time	**1.4±0.16**	**0.08±0.07**	**0.73±0.04**	**0.3±0.2**
	Rule base	37.6±17.3	160±47.5	49.6±17.1	125.1±79.9
OSELM	CR	0.62±0.07	0.68±0.1	0.74±0.13	0.88±0.03
	Rule	50	100	600	35.3±4.16
	Time	1.8±0.2	0.3±0.04	7.1±0.3	1.22±0.13
	Rule base	50	100	100	2118
eClass	CR	**0.69±0.03**	0.76±0.14	0.72±0.3	0.89±0.13
	Rule	14.7±3.13	10.6±3.1	8.8±2.2	12.9±3.1
	Time	4±0.9	0.1±0.02	7.4±2.1	0.43±0.07
	Rule base	132.1±28.2	222.6±64.3	79.4±18.7	154.5±37.5
IT2McFIS	CR	0.87±0.05	0.75±0.13	0.66±0.13	0.89±0.04
	Rule	23.1±2.4	14.5±6.4	8.05±1.9	22.4±3.98
	Time	3.6±0.6	0.19±0.06	0.93±0.4	0.52±0.12
	Rule base	94.6±9.7	122±51.2	34.2±7.6	114.85

IV. EXPERIMENTATION

The efficacy of eT2Class is numerically validated by benefiting from four data streams, featuring dynamic and evolving properties. We encompass four data streams devised in [25] termed drifting Gaussian, rotating checkerboard, shifting hyper-plane problems. The shifting hyper-plane data stream poses the SEA problem [26], which is customized to incorporate the class imbalanced problem in [25]. Apart from that, the semi artificial problem, namely a car study case, is amalgamated, in which it stems from the UCI machine learning repository (http://www.ics.uci.edu/mlearn/MLRepository.html) and is customized in [27] to include the non-stationary component. Note that these problems characterize various concept drifts, involving slow, abrupt, cyclic, gradual time-varying components, which are relevant to test eT2Class. Our experimental procedure and the characteristics of the data streams are akin to [25], while our computational resources rely on an Intel (R) core (TM) i7-2600 CPU @3.4 GHz processor and 8 GB memory. Conversely, eT2Class is benchmarked with its counterparts, namely eClass of [4], pClass of [6], OS-ELM of [29], IT2McFIS of [31], where the predefined parameters of these classifiers are determined with respect to the rule of thumb in their original publications. All consolidated classifiers are assessed in four evaluation standpoints: Classification Rate (CR), number of rule, execution time, rule base parameter. Average numerical results of benchmarked classifiers are summarized in Table 1.

Obviously, eT2Class can outperform other algorithms in producing more reliable or comparable classification rates by evolving more frugal fuzzy rules. On the other hand, the interval type-2 fuzzy system does augment the computational burden, however, the difference is trivial and the advantage of the interval type-2 fuzzy rule does outweigh its deficiency.

V. CONCLUSION

To address the drawback of existing classifiers, a novel evolving interval type-2 fuzzy system, namely Evolving Type-2 Classifier (eT2Class) is proposed in this paper. eT2Class adopts an open rule base principle, where the rule base can be automatically evolved from data streams in the single-pass and local learning modes. The viability of eT2Class has been rigorously assessed by four dynamic and evolving data streams and been compared with the state-of-the art classifiers, where eT2Class demonstrates the most encouraging learning performances in attaining a tradeoff between accuracy and complexity. Our future work will focus on development of the interval type-2 extreme learning machine.

ACKNOWLEDGEMENTS

The work presented in this paper is partly supported by the Australian Research Council (ARC) under Discovery Projects DP110103733 and DP140101366. We thank to Dr.Katrick Subramanian (NTU, Singapore) for sharing the source code of IT2McFIS.

REFERENCES

[1] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning," *Information Sciences.*, vol. 8, no. 3, pp. 199–249, 1975.

[2] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems, *IEEE Transactions on Fuzzy Syst.*, vol. 7, no. 6, pp. 643–658, (1999).

[3] J. M. Mendel and R. I. John, "Type-2 fuzzy sets made simple," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 117–127, (2002).

[4] P. Angelov and D. Filev, "An approach to online identification of Takagi-Sugeno fuzzy models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*: *Cybernetics*, vol. 34, pp. 484-498 (2004).

[5] M.Pratama, et al, "GENFIS: Towards an effective localist network", *IEEE Transactions on Fuzzy Systems*, on line and in press, (2013).

[6] M.Pratama, S.Anavatti, E.Lughofer, pClas, " An Effective Classifier to Streaming Examples", *IEEE Transactions on Fuzzy Systems,* DOI:10.1109/TFUZZ.2014.2312983 (2014).

[7]M.Pratama, M-J.Er, S.Anavatti, E.Lughofer,N.Wang, I.Arifin, " A Novel Meta-Cognitive-based Scaffolding Classifier to Sequential Non-stationary Classification Problems", *in proceeding of IEEE International Conference on Fuzzy Systems*, Beijing , pp.369-376, 2014.

[8] A.Bouchachia, C.Vanaret," GT2FC: An Online Growing Interval Type-2 Self-Learning Fuzzy Classifier", *IEEE Transactions on Fuzzy Systems,* Vol.22, no.4, pp.999-1018, (2014).

[9] 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).

[10] C. F. Juang and Y. W. Tsao, "A type-2 self-organizing neural fuzzy system and its FPGA implementation," *IEEE Transactions on Systems, Man, and Cybernetics, Part-B: Cybernetics*, vol. 38, no. 6, pp. 1537–1548, (2008).

[11] R. H. Abiyev and O. Kaynak, "Type-2 fuzzy neural structure for identification and control of time-varying plants," *IEEE Transactions on Industrial Electronics* , vol. 57, no. 12, pp. 4147–4159,(2010).

[12] Y. Y. Lin, J. Y. Chang, and C. T. Lin, "A TSK-type based self-evolving compensatory interval type-2 fuzzy neural network (TSCIT2FNN) and its applications," *IEEE Transactions on Industrial Electronics.*, vol. 61, no. 1, pp. 447–459, (2014).

[13] Y-Y.Lin, S-H.Liao, J-Y.Chang, C-T.Lin, " Simplified Interval Type-2 Fuzzy Neural Networks", *IEEE Transactions on Neural Networks and Learning Systems,* Vol.25, no.5, pp.959-969, (2014).

[14] L. M. Silva, L. A. Alexandre, and J. Marques de Sa, "Neural network classification: Maximizing zero-error density," in S. Singh et al. (eds.), *Lecture Notes in Computer Science: Pattern Recognition and Data Mining,* vol. 3686, pp. 127-135 (2005).

[15] L.Wang, H-B.Ji, Y.Jin, " Fuzzy Passive-Aggressive Classification: A Robust and Efficient Algorithm for Online Classification Problems", *Information Sciences*, Vol.220,pp.46-63, (2013).

[16] K.Subramanian, S.Suresh, N.Sundararajan, " A metacognitive neuro-fuzzy inference system (mcfis) for sequential classification problems". *IEEE Transactions on Fuzzy Systems*, Vol.21, no.6, pp.1080–1095, (2013).

[17] E. Lughofer*, Evolving Fuzzy Systems --- Methodologies, Advanced Concepts and Applications*, Springer, Heidelberg, (2011).

[18] C-F.Juang, C-Y.Chen, "Data-Driven Interval Type-2 Neural Fuzzy System With High Learning Accuracy and Improved Model Interpretability", *IEEE Transactions on Cybernetics,* Vol.43, no.6, pp.1781-1795, (2013).

[19] S.W.Tung, C.Quek, C.Guan, "eT2FIS: An Evolving Type-2 Neural Fuzzy Inference System", *Information Sciences,* vol.220, pp.124-148, (2013).

[20] D.Wu, J.M.Mendel, "A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets", *Information Sciences,* Vol.178, pp.381-402, (2008).

[21] C.T.Lin, C.S.G.Lee, "Reinforcement structure/parameter learning for neural-network based fuzzy logic control systems," *IEEE Transactions on Fuzzy Systems.*, vol. 2, pp.46–63, (1994).

[22] M.Pratama, S.Anavatti, P.Angelov, E.Lughofer, PANFIS: A Novel Incremental Learning, *IEEE Transactions on Neural Networks and Learning Systems*, Vol.25, no.1, pp.55-68,(2014).

[23] R. R. Yager, "A model of participatory learning," *IEEE Transaction on Systems., Man, Cybernetics.*, vol. 20, no. 5, pp. 1229–1234,(1990).

[24] Y.Xu, K.W.Wong, C.S.Leung, " Generalized Recursive Least Square to The Training of Neural Network", *IEEE Transactions on Neural Networks*, Vol.17,no.1, (2006)

[25]G.Ditzler, R.Polikar, "Incremental Learning of Concept Drift from Streaming Imbalanced Data", *IEEE Transactions on Knowledge and Data Engineering,*Vol.25(10),pp.2283-2301, (2013).

[26] W.N.Street, Y.Kim, "A streaming ensemble algorithm SEA for large-scale classification", in the proceeding of 7th ACM SIGKDD, pp 377-382, (2001).

[27] L.LMinku, X.Yao, " DDD: A New Ensemble Approach for Dealing with Drifts", *IEEE Transactions on Knowledge and Data Engineering,* Vol.24(4), (2012).

[28]J.M.Mendel, " Type-2 Fuzzy Sets and Systems: An Overview", *IEEE Computational Intelligence Magazine*, vol.2(1), pp.20-29,(2007).

[29] N.-Y. Liang, et al, " A Fast and Accurate On-line Sequential Learning Algorithm for Feedforward Networks", *IEEE Transactions on Neural Networks,* vol.17(6), pp.1411-1423, (2006)

[30] H. Bustince, J. Fernandez, H. Hagras, F. Herrera, Interval Type-2 Fuzzy Sets are generalization of Interval-Valued Fuzzy Sets: Towards a Wider view on their relationship. *IEEE Transactions on Fuzzy Systems,* (2015) doi: 10.1109/TFUZZ.2014.2362149, in press (2015)

[31] K.Subramanian, A.K.Das, S.Sundaram, S.Ramasamy, " A Meta-Cognitive Interval Type-2 Fuzzy Inference System and Its Projection-based Learning Algorithm", *Evolving Systems*, DOI:10.1007/s12530-013-9102-9