Manuscript ID: TFS-2014-0178

Title : Evolving Type-2 Fuzzy Classifier

Authors : Mahardhika Pratama, Jie Lu Guangquan Zhang

Dear Editor-in-Chief,

We would first like to express our sincere gratitude to you, because our paper can be eventually published in your transaction. We are really glad that we can contribute our paper again to the IEEE TFS. We also thank the Associate Editor and all reviewers for their help in revising the paper. Because no comments and corrections from the reviewer are found in the decision letter and the author centre, we submit the latest version of our paper as the final file.

Your assistance is much appreciated

Yours sincerely

Dr. Mahardhika Pratama

Corresponding Author

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

Evolving Type-2 Fuzzy Classifier

Mahardhika Pratama, *Member, IEEE*, Jie Lu, *Senior Member, IEEE*, Guangquan Zhang

Centre of Quantum Computation and Intelligent System, University of Technology, Sydney, Australia, email: pratama@ieee.org,jie.lu@uts.edu.au,guangquan.zhang@uts.edu.au

Abstract— Evolving Fuzzy Classifiers (EFCs) have achieved immense success in dealing with non-stationary data streams because of their flexible characteristics. Nonetheless, most real-world data streams feature highly uncertain characteristics, which cannot be handled by the type-1 EFC. A novel interval type-2 fuzzy classifier, namely Evolving Type-2 Classifier (eT2Class), is proposed in this paper, which constructs an evolving working principle in the framework of interval type-2 fuzzy system. The eT2Class commences its learning process from scratch with an empty or initially trained rule base and its fuzzy rules can be automatically grown, pruned, recalled and merged on the fly referring to summarization power and generalization power of data streams. In addition, the eT2Class is driven by a generalized interval type-2 fuzzy rule, where the premise part is composed of the multivariate Gaussian function with uncertain non-diagonal covariance matrix, while employing a subset of the non-linear Chebyshev polynomial as the rule consequents. The efficacy of the eT2Class has been rigorously assessed by numerous real-world and artificial study cases, benchmarked against state-of-the art classifiers, and validated through various statistical tests. Our numerical results demonstrate that the eT2Class produces more reliable classification rates while retaining more compact and parsimonious rule base than state-of-the art EFCs recently published in the literature.

Index Terms—evolving fuzzy system, fuzzy neural networks, type-2 fuzzy system, sequential learning

I. INTRODUCTION

A classifier, which is capable of self-organizing its knowledge from streaming data, is highly desired for coping with real world data streams [1],[2],[46]. Additionally, the classifier should be scalable to cultivate very large data streams, because large volumes of data are continuously generated from sensors, the Internet, etc., at a high incoming rate in today's real-world problems. A retraining process, as done in traditional classifiers [3], is inappropriate for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

overcoming rapidly changing environments because it imposes the considerable computational and memory burden.

Evolving Fuzzy System (EFS) [4]-[9] is a research area of growing interest for learning from data streams. Even so, most EFSs are generally built upon the type-1 fuzzy system, which possesses a crisp and certain membership [47]. The type-1 fuzzy system is, however, not robust to cope with uncertainties in the data representations which can be viewed in the inexact, inaccurate and uncertain characteristics of real world data streams. Uncertainties in the data representations result from disagreements in expert knowledge, noisy measurement, and noisy data. The idea of the type-2 fuzzy system proposed by Zadeh in [10] is an answer to uncertainties in the data representations [11],[12], because it features a fuzzy membership triggered by the so-called fuzzy-fuzzy set. Nonetheless, the viability of the type-2 fuzzy system is in practice hindered by its demanding computational cost, attributed by the type reduction procedure from type-2 to type-1. Therefore, this issue has led to an interval type-2 fuzzy system, which presents a simplified version of the pure type-2 fuzzy system. The interval type-2 fuzzy system can be seen as a special case of the type-2 fuzzy system, which assumes the secondary grade of the type-2 membership functions to be unity to mitigate the computational burden. This idea has been brought into EFS in [13]-[15].

The underlying disadvantage of interval type-2 EFSs in [13]-[15], however, is the over-dependency on the Karnik and Mendel (KM) iterative procedure, which is not in line with the spirit of incremental learning. To correct this drawback, the notion of q coefficient was put forward in [16], in which it orchestrates the proportion of upper and lower bounds in the final output to perform the type reduction. The idea of q coefficient was implemented in [17], and was extended in [18], where in addition to the uncertainty of the premise part, the interval uncertainty

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

is applied in the consequent part. This concept, therefore, leads to a proposal of *ql* and *qr* coefficients. The so-called meta-cognitive interval type-2 neuro-fuzzy system was developed in [19], in which the interval type-2 version of McFIS in [20] is introduced. Although some seminal works in the interval type-2 EFSs can be observed in the literature [13]-[20], a more in-depth study in this area is highly demanded for three reasons: 1) these works still rely on the firing strength-based clustering strategy, which is prone to outliers [49]. 2) The interval type-2 EFSs in [13]-[14], and [16]-[18] suffer from the absence of the rule base simplification strategy in both fuzzy rule and input feature levels. This shortcoming often imposes a prohibitive rule base to be evolved, because the fuzzy rules once added to the rule base cannot be pruned. As a matter of fact, the fuzzy rules can be superfluous as a result of incorrectly incorporating outliers as fuzzy rules and/or a redundancy issue due to lying on a significantly overlapping region. The fuzzy rules can also get outdated owing to changing data distributions. On the other hand, although the interval type-2 EFSs in [18],[30],[32] are equipped by the rule merging scenario to resolve the redundancy issue, they do not take into account the issue of superfluous and obsolete fuzzy rules, whereas in [19] the redundancy aspect is unexplored; 3) Interval type-2 EFCs have not been adequately studied, and many existing EFSs largely focus on regression problems.

A seminal evolving interval type-2 fuzzy classifier, namely Evolving Type-2 Classifier (eT2Class), is presented in this paper. It is evolving in the sense that it can carry out automatic knowledge building, and can forget superfluous knowledge, while at the same time being computationally efficient because all learning procedures are committed in the single-pass learning mode. The algorithmic development of the eT2Class involves six salient learning components, as follows: 1) the fuzzy inference scheme of the eT2Class is reliant on a generalized version of the interval type-2 Gaussian rule with uncertain Standard Deviations (SDs), where the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

rule premise part is underpinned by the multivariable Gaussian function with uncertain non-diagonal covariance matrix, while the rule consequent is compiled by a subset of the Chebyshev polynomial, expanding the degree of freedom of the Takagi Sugeno Kang (TSK) rule consequent. 2) The fuzzy rules are autonomously extracted from data streams by two rule growing cursors: Type-2 Datum Significance (T2DS); Type-2 Data Quality (T2DQ) methods, which are extended from its type-1 model in [8] to suit the type-2 fuzzy rule. 3) A new fuzzy rule is initialized with the class overlapping criterion to warrant a strategic influence zone of the new fuzzy rule. This strategy is to avoid a possible class overlapping phenomenon possibly degenerating the classifier's generalization [20]. 4) Obsolete and superfluous fuzzy rules can be captured and in turn pruned by two rule pruning scenarios: Type-2 Extended Rule Significance (T2ERS); Type-2 Potential+ (T2P+) concepts, in which their original versions in [8] are redefined to deal with the type-2 fuzzy rule. The T2P+ method is also useful for carrying out the rule recall mechanism, where the fuzzy rule, deactivated by this method in earlier training episodes, can be recalled in the future to handle a recurring or cyclic concept drift. 5) To boost the interpretability of rule semantics, the eT2Class is endued by a novel rule merging scenario using multi-faceted merging criteria: the vector similarity measure in [21], [22]; and the blow-up check. 6) Lastly, the eT2Class makes use of the *ql* and *qr* coefficients to implement the type reduction mechanism, where a novel adaptation scheme inspired by the Zero Error Density Maximization (Z-EDM) method [23] to refine the *ql* and *qr* coefficients is proposed. Furthermore, the convergence of this method is mathematically proven in this paper via the Lyapunov stability criterion. On the other hand, the rule outputs are fine-tuned via a local learning scenario of the Fuzzily Weighted Generalized Least Square (FWGRLS) method [24].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

The major contributions are as follows: 1) this paper proposes a generalized interval type-2 fuzzy rule with uncertain SDs, featuring the multivariate Gaussian function in the premise part and the functional link-based Chebyshev polynomial in the consequent part. 2) Several novel learning technologies are proposed in the eT2Class such as rule growing, merging, pruning scenarios and the adaptation of *ql* and *qr*. 3) The efficacy of the proposed classifier has been numerically validated with the use of numerous numerical problems, benchmarks with its counterparts and statistical tests, which confirm the effectiveness of the eT2Class in achieving a trade-off between accuracy and complexity.

The remainder of this paper is organized as follows: Section II outlines the eT2Class's inference scheme. Section III details the algorithmic development of the eT2Class. Section IV elaborates the sensitivity analysis of the predefined parameters. Section V discusses the empirical studies. Section VI analyses the conceptual comparisons of eT2Class with other prominent interval type-2 fuzzy classifiers. Conclusions are drawn in Section VII.

II. Network Architecture of ET2Class

The salient learning property of the eT2Class is shown by its fuzzy rule type, where the multivariate Gaussian function is assembled in the rule premise and the functional link-based Chebyshev polynomial is mounted in the rule consequent. The fuzzy rule of the eT2Class is expressed as follows:

$$R_i: \textbf{IF}\ \ X \textit{ is close to } \tilde{R}_i \ \textbf{Then}\ \ y_i^{\,o} = x_e \tilde{\Omega}_i$$

where $\tilde{R}_i = [\underline{R_i}, \overline{R_i}]$ denotes a multidimensional kernel compiled by the multidimensional Gaussian function with uncertain non-diagonal covariance matrix, which can be written as follows:

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

where $C_i \in \Re^{1\times u}$ stands for the centroid of the Gaussian function of the *i-th* rule and *u* designates the number of input dimensions. $\tilde{\Sigma}_i^{-1} \in \Re^{u\times u}$ labels a non-diagonal inverse covariance matrix,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

whose elements exhibit the inter-relation between input variables, which then govern the orientation of the ellipsoidal clusters. Specifically, the interval type-2 fuzzy rule with fixed mean and uncertain SD is used in the eT2Class with the use of the uncertain inverse covariance matrix $\tilde{\Sigma}_i^{-1} = [\Sigma_{i,1}^{-1}, \Sigma_{i,2}^{-1}]$. The merit of this rule type can be perceived in its scale-invariant feature and its ability to hamper information loss of input feature interactions. By extension, this sort of fuzzy rule guarantees a proper cluster shape, particularly when the streaming data are not distributed in the main axes, thereby being able to diminish the required number of fuzzy rules. We use the Gaussian function here because it can avoid undefined states due to its infinite support and can render the smooth approximation of the local data space owing to its steady differentiable merit.

In the rule consequent part, y_i^o denotes the regression output of the *o-th* class in the *i-th* rule, and $\tilde{\Omega}_i$ labels a weight vector, which can be formulated as $\tilde{\Omega}_i \in \Re^{(2u+1)\times m}$, where *m* specifies the output dimensionality. In what follows, $\tilde{\Omega}_i$ is defined by an interval set as $\tilde{\Omega}_i = [\Omega_l^{i,o}, \Omega_r^{i,o}]$, where $\Omega_l^{i,o} = [w_1^{i,o,l}, ..., w_{2u+1}^{i,o,l}]$ $\Omega_r^{i,o} = [w_1^{i,o,r}, ..., w_{2u+1}^{i,o,r}]$, and $x_e \in \Re^{1\times(2u+1)}$ formulates an expanded input vector as a result of a non-linear mapping based on the Chebyshev polynomial up to the second order. It is worth noting that the eT2Class is equipped by the Chebyshev polynomial [25], because the standard form of the TSK rule consequent [26] generates a linear hyper-plane, which cannot properly reveal a local approximation trait. This drawback is rectified by expanding the degree of freedom of the rule consequent via a non-linear mapping of the Chebyshev polynomial. The mathematical expression of the Chebyshev polynomial is given as follows:

$$T_{n+1}(x) = 2x_j T_n(x_j) - T_{n-1}(x_j) \tag{2}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

with $T_0(x_j)=1$, $T_1(x_j)=x_j$, $T_2(x_j)=2x_j^{\,2}-1$. Suppose X is a 2-D input pattern $X=[x_1,x_2]$. The expanded input vector becomes $x_e=[1,x_1,T_2(x_1),x_2,T_2(x_2)]$. Note that we include the term 1 in this case to include the intercept of the rule consequent as the case in the standard form of the TSK fuzzy rules (otherwise, all consequent hyper-planes may go through the origin, which will lead to untypical gradients).

The main bottleneck of the multivariate Gaussian function to be integrated in the interval type-2 fuzzy inference scheme is that it does not have the fuzzy set representation or membership function, which is solicited to infer a crisp output. Hence, the Footprint of Uncertainty (FOU), which is a unique attribute of the interval type-2 fuzzy system, cannot be explicitly exhibited by the fuzzy rule [11]. In addition, its rule semantic is less interpretable due to the absence of atomic clauses in the rule antecedent. To remedy this bottleneck, we can benefit from our in-depth investigation in [8], which conveys two avenues to extract a fuzzy set representation of the multivariate Gaussian function. The second method is used in the eT2Class, since the first one necessitates quantifying the eigenvalue and eigenvector in every training step, which inevitably retards the training process. Even so, the disadvantage of the second approach is that it designates an inaccurate fuzzy set representation, should the ellipsoidal rule be rotated around 45 degrees. The fuzzy region of the multivariate Gaussian function can be defined as follows:

$$\tilde{\sigma}_i=\frac{r}{\sqrt{\tilde{\Sigma}_{ii}^{\,1,2}}},\ \tilde{\sigma}_i\in[\sigma^1{}_i,\sigma^2{}_i],\ \tilde{\Sigma}_i\in[\Sigma^1{}_i,\Sigma^2{}_i] \tag{3}$$

where Σ_{ii} stands for the diagonal elements of the inverse covariance matrix and r defines a Mahalanobis distance between the datum and the *i-th* cluster. Note that the fuzzy set center is equivalent to the center of the multivariate Gaussian function. Executing (3) allows us to proceed to the interval type-2 fuzzy inference scheme, because we presume upper and lower multivariate Gaussian functions to be distinct in higher dimensional space with a specific uncertainty region

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

separating them. Formally speaking, the interval type-2 inference scheme can be triggered according to [11] as follows:

$$\tilde{\mu}_{i,j} = \exp(-(\frac{x_j - c_{i,j}}{\tilde{\sigma}_{i.j}})^2),\ \tilde{\sigma}_{i,j} \in [\sigma^1{}_{i,j}, \sigma^2{}_{i,j}] \tag{4}$$

thus being able to form upper and lower MFs as follows:

$$\overline{\mu}_{i,j} = N(c_{i,j}, \sigma^1{}_{i,j}; x_j),\ \underline{\mu}_{i,j} = N(c_{i,j}, \sigma^2{}_{i,j}; x_j) \tag{5}$$

where $\sigma^1{}_{i,j} > \sigma^2{}_{i,j}$. We produce the spatial firing strength of *i-th* rule via the product *t-norm* operator, which in turn inflicts $\tilde{R}_i = [\underline{R_i}, \overline{R_i}]$ as follows:

$$\underline{R_i} = \prod_{j=1}^{u} \underline{\mu_{i,j}},\ \overline{R_i} = \prod_{j=1}^{u} \overline{\mu_{i,j}} \tag{6}$$

The design factors $[q_l, q_r]$ are employed to perform the type reduction mechanism, which transforms the type-2 output sets to the type-1 fuzzy sets - called type reduced sets. This reduction method expedites the training process compared to the classical K-M iterative procedure, which requires an expensive iterative procedure to elicit the L and R end points. The coefficients $[q_l, q_r]$ are to be adaptively adjusted by the Z-EDM method, thereby being able to autonomously control the proportion of the upper and lower outputs $[y_l, y_r]$ in respect to the uncertainty degree contained in the data streams.

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

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

where *P* is the number of existing fuzzy rules. The outputs $[y_l, y_r]$ pinpoint the type reduced sets, which results in a crisp output. If the MIMO architecture is in charge to infer the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

classification result [2], the final output is assembled with the help of the maximum operator as follows:

$$y_o = y_{l,o} + y_{r,o}, y = \max_{o=1,..,m}(\hat{y}_o) \quad (9)$$

It is worth noting that other types of classifier architectures, namely one against all or one against all, can be applied to output the classification decision of the eT2Class. The MIMO architecture is used to evolve the classification decision in this paper, because it is omnipresent in the literature, thus potentially allowing for fair comparison with its counterparts. This facet has been verified in [8], where an evolving classifier, namely pClass, is tested by using the three classifier's architectures: MIMO, one against all, one against one. eT2Class can be seen as an extension of pClass, thus being expected to accommodate the three architectures equally well as pClass. The network topology of the eT2Class is visualized by Fig.1, whereas Fig.2 visualizes the Gaussian interval type-2 fuzzy set with uncertain SD.

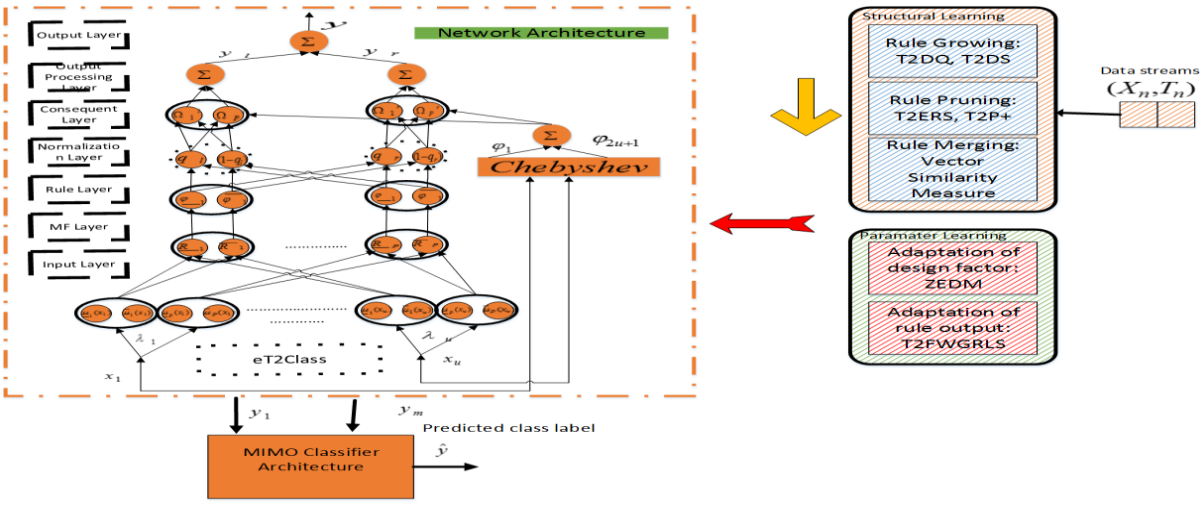

Fig.1 Learning Architecture of ET2Class

III. Rule Base Management of ET2Class

The learning policies of the eT2Class are divided into six sub-sections, as detailed below. Algorithm 1 exhibits an overview of the eT2Class's learning policy.

Rule Generation Mechanism: The contribution of data streams is examined by two rule growing scenarios, which aim to enumerate the statistical contribution of data stream and the data density.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

This notion is equivalent to our past work in [8], which is reconfigured for the type-2 fuzzy system in this paper. The first strategy is to utilize the T2DS method, which is derived in the same way of the Datum Significance (DS) method in [8]. Generally speaking, the statistical contribution of the hypothetical cluster (P+1st) can be defined as follows:

$$DS_n = \lim_{N\to\infty}\left(\frac{\sum_{n=1}^{N} R_{P+1}/N}{\sum_{i=1}^{P+1}\sum_{n=1}^{N} R_i/N} + \frac{\sum_{n=1}^{N} \overline{R_{P+1}}/N}{\sum_{i=1}^{P+1}\sum_{n=1}^{N} \overline{R_i}/N}\right)/2 \tag{10}$$

where *N* labels the number of training observations. Nevertheless, this expression is cumbersome in the online environment, because it entails having complete access to already seen training patterns. To this end, we can simplify this mathematical expression as [9], as follows*:*

$$DS_n = \frac{1}{2S(X)}\left(\int_X \exp(-(X_n - C_{P+1})^T \Sigma_{p+1,1}^{-1}(X_n - C_{p+1}) + \exp(-(X_n - C_{p+1})\Sigma_{p+1,2}^{-1}(X_n - C_{p+1}))\right)dx \tag{11}$$

where $S(X)$ denotes a size of range X, which can be obtained from $S(X) = \int_X 1dx$ with the assumption that the sampling density function follows a uniform distribution $p(x) = \frac{1}{S(x)}$, whereas $C_{p+1}, \Sigma_{p+1}^{-1}$ refers to the hypothetical new rule *P+1*. In practice, it is impractical to fix $S(X)$ a priori because a true data distribution is unknown. Therefore, it is substituted by a total contribution of existing fuzzy rules for simplicity. Furthermore, (11) can be solved via *P+1*-fold numerical integration for any arbitrary density function $l(x)$, thus expressing the final formula of the T2DS method as follows:

$$DS_n = \frac{1}{2}\left(\frac{V_{P+1}^{1}}{\sum_{i=1}^{P+1} V_{P+1}^{1}} + \frac{V_{P+1}^{2}}{\sum_{i=1}^{P+1} V_{P+1}^{2}}\right) \tag{12}$$

where the volume of the *P+1-th* multivariate Gaussian rule can be simply enumerated by means of the *det* operator. If a more precise estimation is needed, it can be computed according to [8]. To omit a predefined threshold, if the hypothetical cluster possesses a higher volume than those

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

of existing clusters, it inevitably offers a substantial statistical contribution to be a new rule as follows:

$$(V_{P+1}^{\ 1} + V_{P+1}^{\ 2}) > \max_{i=1,..,P} (V_i^{\ 1} + V_i^{\ 2}) \tag{13}$$

Algorithm 1: Pseudo code of ET2Class

Define: Input attributes and Desired class labels: $(X_n, T_n) = (x_1,\ldots,x_u, t_1,..,t_m)$

Predefined Thresholds: $\rho_2 = 0.01, \rho_3 = 0.8, \rho_4 = 1.1, \rho_5 = 0.9, \varpi = 10^{-15}$

/*Phase 1: Rule Growing and Adaptation of the Fuzzy Rule Premise /*

For i=1 to P do

Compute the posterior probabilities of the fuzzy rules (22)

Update the T2DQ method for all rules (15)

Compute the volume of existing rules using *det* operation

End For

Determine the winning rule $win = \arg\max_{i=1,\ldots,P} \bar{P}(R_i|X)$

Compute the T2DS method (12) and update the T2P+ method for P* rules(32)

IF (13) and (16) Then Rule Growing=true

IF $\max_{i^*=1,..,P^*} (\chi_{i^*}) > \max_{i=1,..,P+1} (DQ_i)$ ***Then***

Activate rule recal mechanism (33)

Else IF

IF $(\underline{R_{win}} + \overline{R_{win}})/2 \geq \rho_a$ ***Then***

Compute the data quality per class method (17)

IF $\max_{o=1,..,m} (DQ_o) \neq true\_class\_label$ ***Then***

Initialize the new fuzzy rule as (17)

Else IF

Initialize the new fuzzy rule as (18)

End IF

Else IF

Initialize the new fuzzy rule as (19)

End IF

Assign the new rule consequent as (21)

Else IF

Adjust the rule premise (25)-27

End IF

/*Phase 2: Rule Pruning and Merging Strategy /*

IF *Rule Growing=False* ***Then***

For i=1 to P do

Enumerate the T2ERS and T2P+ method (30) and (32)

IF $ERS_i^{\ n} \leq mean(ERS_i^{\ n}) - std(ERS_i^{\ n})$ ***Then***

Prune the fuzzy rules

End IF

IF $\chi_i^{\ n} \leq mean(\chi_i^{\ n}) - std(\chi_i^{\ n})$ ***Then***

Deactivate the fuzzy rules subject to the rule recall mechanism P*=P*+1

End For

For i=1 to P do

For j=1 to U do

Compute the shape-based and proximity-based similarity measures(36,37)

Quantify the vector similarity measure (34)

End For

IF (40),(41) Then

Coalesce the fuzzy rules (42)-(48)

End IF

End For

End IF

/*Phase 3: Adaptation of Rule Consequents and Design Factors /*

For i=1 to P do

Adjust the fuzzy rule consequents and design factors (49)-(52) and (54)

End For

The T2DQ method, having its root in the DQ method of [24], is embedded as the second rule growing cursor, which is tailored according to the specification of interval type-2 fuzzy rules. The crux of this method is to delve the spatial proximity of a datum with all preceding samples without maintaining previous training stimuli in the memory as follows:

$$DQ_n = \frac{1}{2}\sqrt{\frac{1}{(1+\frac{\sum_{n=1}^{n-1} DQ_n \sum_{j=1}^{u} (x_j^{\ n} - \underline{c}_j P+1)}{\sum_{n=1}^{n-1} DQ_n})} + \frac{1}{(1+\frac{\sum_{n=1}^{n-1} DQ_n \sum_{j=1}^{u} (x_j^{\ n} - \bar{c}_j P+1)}{\sum_{n=1}^{n-1} DQ_n})}} \tag{14}$$

Since an identical centroid is used in both upper and lower fuzzy rules, we can define DQ_n via a recursive computation as follows:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

$$DQ_N = \sqrt{\frac{U_n}{U_n(1+b_n)-2h_n+g_n}} \tag{15}$$

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

Note that the recursive parameters are initialized as zero prior to commence the training process. The underlying contribution of the data stream is appealing for the eT2Class given that it is either densely populated by most other samples or posited in the remote region, which is uncharted by existing fuzzy rules. In a nutshell, a new fuzzy rule is added if the data stream complies with one of these two criteria, as follows:

$$DQ_N \geq \max_{i=1,..,P}(DQ_i) \text{ or } DQ_N \leq \min_{i=1,...,P}(DQ_i) \tag{16}$$

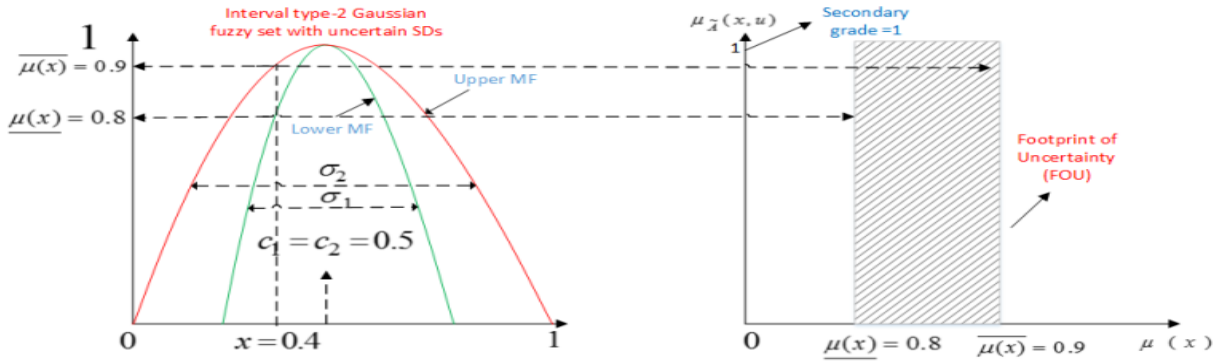

Fig.2 Interval type-2 Gaussian MF

Our DQ method is unlike that of the rule potential of [1], which suffers from an outlier's shortfall [27]. We engage the weighting factor to weight the pairwise distances with the membership values of the previous samples DQ_{N-1}, thus lowering the leverage of the outlier to the following samples. In addition, the condition $DQ_N \leq \min_{i=1,...,P}(DQ_i)$ should be equipped with a rule pruning scenario to forestall an outlier to be included in the rule base. The criterion $DQ_N \geq \max_{i=1,..,P}(DQ_i)$ should be circumspectly consolidated to avert the class overlapping problem, since it is prone to be in close proximity to inter-class clusters.

Fuzzy Rule Initialization: In the EFC, the initialization of a new fuzzy rule plays an important role in partitioning the input space, since it might cause the class overlapping phenomenon,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

where the new fuzzy rule is crafted in an unsafe distance with the inter-class cluster. This issue is unexplored in most interval type-2 EFSs [13]-[19]. In [19],[20], a distance ratio between inter- and intra-class clusters were proposed to prevent the class overlapping phenomenon, but, this work excludes a case of a non-purified cluster, in which a cluster comprises different-class samples. In [28], we devised the quality per class concept, which takes into account the neighbouring relation of the incoming datum and the class labels. In this paper, we enhance our work in [28] to be implemented in the type-2 fuzzy system.

First, we compute the spatial proximity between the newly-created fuzzy rule and the winning rule to envision the likelihood of the class overlapping situation. A remarkable situation is pinpointed by $(\underline{R_{win}} + \overline{R_{win}})/2 \geq \rho_a$, where the new fuzzy rule is deemed too close to the winning rule. ρ_a is a predefined constant that can be statistically represented by the critical value of a χ^2 distribution with Z degrees of freedom and a significance level of α [29], termed as $\chi_p{}^2(\alpha)$. A typical value of α is 5%, and the degree of freedom is represented by the dimensionality of the learning problem, thus resulting in $Z = u$. Therefore, we set $\rho_a = \exp(-\chi_p{}^2(\alpha))$. After that, we need to investigate the class relationship by switching on the quality per class measure to check whether or not the new fuzzy rule is adjacent to inter- or intra-class data clouds. The quality per class method is formulated as follows:

$$DQ_o = \sqrt{\frac{1}{1 + \frac{\sum_{n_o=1}^{N_o} \sum_{j=1}^{u+m} (x_j{}^{n_o} - x_j{}^{N})}{(N_o - 1)}}} = \sqrt{\frac{(N_o - 1)}{(N_o - 1)(ab_n + 1) + cb_n - 2bb_n}} \tag{17}$$

where $ab_n = \sum_{j=1}^{u+m} (x_j{}^{N})^2$, $cb_{no} = cb_{no-1} + \sum_{j=1}^{u+m} (x_j{}^{N_o-1})^2$, $bb_{no} = \sum_{j=1}^{u+m} x_j{}^{N} d_j{}^{no}$, $d_{no} = d_{no-1} + x_{N_o-1}$ and N_o denotes the number of samples falling in the *o-th* class. Meanwhile, $x_j{}^{N}$ respectively stands for the latest

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

incoming datum of the *o-th* class and x_j^{no} denotes the streaming data falling to the *o-th* class. The class overlapping situation is indicated by $\max_{o=1,..,m}(DQ_o) \neq true\_class\_label$, where the new fuzzy rule has a more neighboring relationship with the inter-class data clouds. In essence, the class overlapping condition inevitably slackens the classifier's generalization because an over-complex decision boundary will be created to classify an overlapping region belonging to more than one class. To remedy this situation, suppose that *ie* is the nearest inter-class cluster and *ir* denotes the winning intra-class cluster; the new fuzzy rule should be crafted in such a way to dampen the class overlapping effect as follows:

$$c_{P+1,j}^{1,2} = x_j - \rho_2(c_{ie,j}^{1,2} - x_j),\ dist_1^j = \frac{\rho_1}{N_{win}}\left|c_{P+1,j}^{1} - c_{ie,j}^{1}\right|,\ dist_2^j = \frac{2\rho_1}{N_{win}}\left|c_{P+1,j}^{2} - c_{ie,j}^{2}\right|,\ \tilde{\Sigma}_{p+1}^{-1} = (dist_{1,2}^T dist_{1,2})^{-1} \quad (18)$$

where $\rho_1 = r_{ir}^j / r_{ie}^j$ stands for an overlapping factor, steering the shrinkage of the fuzzy set with respect to the distance ratio between the intra- and inter-class clusters. Obviously, this setting of this inverse covariance matrix is plausible to form the new fuzzy region, since ρ_1 will be automatically decreased, when the new fuzzy rule is more similar to the inter-class winning cluster and vice versa. $\rho_2 \in [0.01-0.1]$ is a problem-independent shifting factor, fixed as 0.01 in all our simulations in this paper. In essence, this fuzzy rule initialization strategy alleviates the coverage span of the new cluster and shifts the centroid of the new cluster away from the winning rule, thus being able to minimize the impact of class overlapping. Note that the upper fuzzy set can be simply formed as the standard deviation of N_{win} supports of the winning cluster, whereas the lower fuzzy set can be assigned as the standard deviation of $N_{win}/2$ populations of the winning cluster referring to [30].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

Another circumstance is shown by $\max_{o=1,..,m}(DQ_o) = true\_class\_label$, where the new cluster is surrounded by intra-class data samples. More confident parameters can be allocated in this case because a new fuzzy rule merely incurs a minor risk of class overlapping as follows:

$$c_{P+1,j}^{1,2} = x_j + \rho_2(c_{ir,j}^{1,2} - c_{ie,j}^{1,2}),\ dist_1^{\,j} = \frac{\rho_1}{N_{win}}\left|c_{P+1,j}^{\,1} - c_{ir,j}^{\,1}\right|,\ dist_2^{\,j} = \frac{2\rho_1}{N_{win}}\left|c_{P+1,j}^{\,2} - c_{ir,j}^{\,2}\right|\ \tilde{\Sigma}_{p+1}^{-1} = (dist_{1,2}^{\,T} dist_{1,2})^{-1} \quad (19)$$

One can argue that the new cluster and the intra-class cluster can be significantly overlapping in the future after obtaining the adaptation of rule premise (25)-(27). It is worth stressing that this situation is unlikely to deteriorate the classifier's generalization, since the decision boundary in the output space is not substantially affected. Moreover, this situation can be resolved by the rule merging scenario in section III.5, where it coalesces redundant clusters into a single cluster.

If the newly-added rule is situated in a remote area to the winning cluster $R_{win} < \rho_a$, the new cluster is tailored as follows:

$$c_{P+1,j}^{1,2} = x_j,\ dist_1^{\,j} = \frac{\rho_1}{N_{win}}\left|x_j - c_{ir,j}^{\,1}\right|,\ dist_2^{\,j} = \frac{2\rho_1}{N_{win}}\left|x_j - c_{ir,j}^{\,2}\right|\ \tilde{\Sigma}_{p+1}^{-1} = (dist_{1,2}^{\,T} dist_{1,2})^{-1} \quad (20)$$

In this regard, the fuzzy rule initialization scenario is negligible to the stability of other local fuzzy sub-models, because it occupies an uncharted fuzzy region.

On the other hand, the rule consequent and the output covariance matrix are set as follows:

$$\Omega_{P+1}^{\,l,r} = \Omega_{win}^{\,l,r}\ \Psi_{P+1}^{\,l,r} = \omega I \quad (21)$$

where ω denotes a large positive constant. The output covariance matrix Ψ_{P+1} is allocated as a positive definite matrix to rapidly emulate a real solution as produced by a batched learning solution [31]. The rule output is crafted as the winning rule output, since the winning rule snapshots a pertinent data trend to a new rule, thereby expediting the convergence.

The category selection mechanism benefits from the Bayesian concept, which selects a category having a maximum posterior probability $win = \arg\max_{i=1,...,P} \widehat{P}(R_i|X)$ as a winning rule. The merit of this approach is capable of handling in the probabilistic standpoint with the use of prior

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

probability formula a situation in which several candidates of winning clusters are located in the similar proximity. The posterior, prior probabilities and likelihood functions are respectively defined as follows:

$$\hat{P}(R_i|X) = \frac{1}{2}\left(\frac{\hat{p}(X|R_i)^1 \hat{P}(R_i)}{\sum_{i=1}^{P} \hat{p}(X|R_i)^1 \hat{P}(R_i)} + \frac{\hat{p}(X|R_i)^2 \hat{P}(R_i)}{\sum_{i=1}^{P} \hat{p}(X|R_i)^2 \hat{P}(R_i)}\right) \tag{22}$$

$$\hat{P}(X|R_i)^{1,2} = \frac{1}{(2\pi)^{1/2} V_{i,1,2}^{1/2}} \exp(-(X - C_i^{1,2})\Sigma_{i,1,2}^{-1}(X - C_i^{1,2})^T) \tag{23}$$

$$\hat{P}(R_i) = \frac{\log(N_i + 1)}{\sum_{i=1}^{P} \log(N_i + 1)} \tag{24}$$

where N_i express the number of data points contained by the *i-th* cluster. The prior probability formula $\hat{P}(R_i)$ is softened from its original definition to allow a newly created category to win the category choice phase more frequently, thus being able to develop its shape.

Rule premise adaptation: If the rule growing criteria in equations (13) and (16) are not satisfied, a datum solely draws a marginal conflict to the existing knowledge base. This condition is easily coped with the rule premise adaptation scenario as follows:

$$C_{win,1,2}{}^{N} = \frac{N_{win}{}^{N-1}}{N_{win}{}^{N-1} + 1} C_{win,1,2}{}^{N-1} + \frac{(X_N - C_{win,1,2}{}^{N-1})}{N_{win}{}^{N-1} + 1} \tag{25}$$

$$\Sigma_{win,1,2}(N)^{-1} = \frac{\Sigma_{win,1,2}(N-1)^{-1}}{1-\alpha} + \frac{\alpha}{1-\alpha} \frac{(\Sigma_{win,1,2}(N-1)^{-1}(X_N - C_{win,1,2}{}^{N-1}))(\Sigma_{win,1,2}(N-1)^{-1}(X_N - C_{win,1,2}{}^{N-1}))^T}{1 + \alpha(X_N - C_{win,1,2}{}^{N-1})\Sigma_{win,1,2}(old)^{-1}(X_N - C_{win,1,2}{}^{N-1})^T} \tag{26}$$

$$N_{win}{}^{N} = N_{win}{}^{N-1} + 1 \tag{27}$$

where $\alpha = 1/(N_{win}{}^{N-1} + 1)$. The rule premise adaptation in (25)-(27) is generalized from the sequential maximum likelihood estimation for a spherical cluster to accommodate the non-axis-parallel ellipsoidal cluster. In addition, (26) is modified to construct a direct update of the inverse covariance matrix to avoid an additional re-inversion step, which can cause numerical instability, after committing the category adjustment.

Rule pruning mechanism: The eT2Class is equipped by two novel pruning mechanisms for the interval type-2 fuzzy system - T2ERS and T2P+ methods. The two methods aim to get rid of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

obsolete and inconsequential fuzzy rules to relive the complexity. These two methods are extended from our past works in [8] to suit to the type-2 fuzzy rule. In essence, the T2ERS method is akin to the T2DS method as the rule growing cursor quantifying the statistical contribution of the existing fuzzy rules. The statistical contribution of the *i-th* fuzzy rule is defined as the total contribution of the rule antecedent and consequent as follows:

$$ERS_i{}^n = \frac{1}{2}|\delta_i|E_i \tag{28}$$

where δ_i is the contribution of the *i-th* rule consequent $\delta_i = \sum_{j=1}^{2u+1}\sum_{o=1}^{m}\Omega_{i,l}{}^{o,j} + \Omega_{i,r}{}^{o,j}$. Conversely, the contribution of the *i-th* rule input is defined as (9) as follows:

$$E_i^n = \frac{1}{S(X)}\left(\int_X \exp(-(X_n - C_i)^T \Sigma_{i,1}{}^{-1}(X_n - C_i) + \exp(-(X_n - C_i)\Sigma_{i,2}{}^{-1}(X_n - C_i))\right)dx \tag{29}$$

where $C_i, \Sigma_{i,1,2}{}^{-1}$ denote the centroid and inverse covariance matrix of the *i-th* rule. By following the same mathematical derivation as the DS method, we arrive at a similar formula (12). We therefore combine this result with (28), thus producing:

$$ERS_i{}^n = \frac{1}{2}\left|\sum_{j=1}^{2u+1}\sum_{o=1}^{m}\Omega_{i,l}{}^{o,j} + \Omega_{i,r}{}^{o,j}\right|\left(\frac{V_i^1}{\sum_{i=1}^{p}V_i^1} + \frac{V_i^2}{\sum_{i=1}^{p}V_i^2}\right) \tag{30}$$

A fuzzy rule is deemed inactive, when encountering $ERS_i{}^n \le mean(ERS_i{}^n) - std(ERS_i{}^n)$. The advantage of the T2ERS method over other rule pruning methods is that it can foresee the contribution of the fuzzy rules in the future. It also involves the contribution of the rule output, which exhibits the fuzzy rule contribution to the overall output.

Another rule pruning method, namely T2P+, is integrated in the eT2Class learning engine to capture the out-dated fuzzy rules which are no longer relevant, to delineate the up-to-date data trends owing to the concept drift. The T2P+ method is defined as follows:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

$$\chi_i^n = \frac{1}{2}\left(\sqrt{\frac{1}{1+\sum_{n=1}^{N-1}\sum_{j=1}^{m+u}\frac{(x_i^j - c_{i,1}^j)^2}{(N-1)}}} + \sqrt{\frac{1}{1+\sum_{n=1}^{N-1}\sum_{j=1}^{m+u}\frac{(x_i^j - c_{i,2}^j)^2}{(N-1)}}}\right) \tag{31}$$

As $c_{i,j}^{\ 1} = c_{i,j}^{\ 2}$, we can derive the final expression of the T2P+ method as follows:

$$\chi_i = \sqrt{\frac{(N-1)\chi_{N-1,i}^{\ 2}}{2\chi_{N-1,i}^{\ 2} + 2\chi_{N-1,i}^{\ 2}\sum_{j=1}^{m+o}(x_{i,j}^{\ N-1} - c_{i,j})^2 + (N-2)}} \tag{32}$$

(31) is derived as with the P+ methods in [8]. The T2P+ method enhances the P+ method in [8] to handle the type-2 fuzzy system. A fuzzy rule is deactivated, if $\chi_i^{\ n} \le mean(\chi_i^{\ n}) - std(\chi_i^{\ n})$ is met. It is worth noting that T2P+ method implies a cluster's density, because it tracks the rule potential overtime, thereby being able to monitor the cluster evolution.

Another appealing property of the T2P+ method is a rule recall mechanism to cope with the recurring concept drift phenomenon. That is, the past data distribution re-appears to fade the current data distribution and it can be found in the weather prediction problem. Therefore, appending a completely new rule without allowing reviving a previously deactivated fuzzy rule, which is compatible to address the cyclic data distribution, does not coincide with the remembering facet of human being. In other words, it can result in the so-called catastrophic forgetting of adaptation history. The fuzzy rules deactivated by the T2P+ method should be subject to the rule recall mechanism in the future, given that their potential is fit to the current data distribution $\max_{i*=1,..,P*}(\chi_{i*}) > \max_{i=1,..,P+1}(DQ_i)$, where P^* stands for the number of fuzzy rules already pruned in the earlier training episodes. The fuzzy rule pruned in earlier training episodes contains previous adaptation history, thus being more suitable to be chosen as a new rule as follows:

$$C_{P+1,1,2} = C_{i*,1,2}, \Sigma_{P+1,1,2}^{-1} = \tilde{\Sigma}_{i*,1,2}^{-1}, \tilde{\Psi}_{P+1}^{l,r} = \tilde{\Psi}_{i*}^{l,r}, \Omega_{P+1}^{l,r} = \Omega_{i*}^{l,r} \tag{33}$$

Nonetheless, the pruned fuzzy rules are limited to solely execute (33) without being involved in other learning scenarios, thereby still alleviating the computational burden. It is worth-stressing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

that the ERS and P+ methods are activated given that no fuzzy rule is added to the rule base. This strategy is to circumvent a new fuzzy rule to be pruned after being appended to the rule base.

Rule merging scenario: The use of the rule merging scenario in the interval type-2 fuzzy neural network was pioneered in [30], [32], which benefits from either the shape-based or distance-based similarity measure without synergizing these two methods simultaneously. On the other hand, most interval type-2 rule merging mechanisms discount the so-called blow up test, thus possibly merging non-homogenous clusters. Such merging scenario causes an inexact representation of two or more non-homogenous local data clouds in the merged cluster and leads to the so-called cluster delamination. To remedy the bottlenecks, the eT2Class rule merging scenario encompasses a multi-faceted strategy: the similarity test; the blow-up check. The similarity measure is undertaken by the so-called vector similarity measure [21], [22], accounting both the shape and proximity of two fuzzy sets in vetting the similarity. On the other hand, the blow-up check scrutinizes the volume of clusters to oversee the possibility of the blow up effect.

The main notion of the vector similarity method is to approximate the similarity of two interval type-2 fuzzy sets based on their shape and distance in one joint formula as follows:

$$s_{v,j}(win,i) = s_{1,j}(win,i) \times s_{2,j}(win,i) \tag{34}$$

where $s_{1,j}(win,i) \in [0,1]$ denotes the shape-based similarity measure and $s_{2,j}(win,i) \in [0,1]$ labels the distance-based similarity measure, while the resultant similarity check is elicited by the product operator. Since shape and distance-based similarity measures are used, $\tilde{win}$ and $\tilde{i}$ fuzzy sets are aligned to quantify the shape-based similarity measure more accurately. The alignment procedure involves moving one or both $\tilde{win}$ and $\tilde{i}$, so that $c_{win,j}^{1,2}$ and $c_{i,j}^{1,2}$ coincide $c_{win,j}^{1,2} = c_{i,j}^{1,2}$. To this end, we apply the extended Jaccard similarity measure with the use of the average cardinality principle as follows:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

$$s_{1,j}(win,i)=\frac{M(\underline{\mu_{win,j}}\cap\underline{\mu_{i,j}})+M(\overline{\mu_{win,j}}\cap\overline{\mu_{i,j}})}{M(\underline{\mu_{win,j}}\cup\underline{\mu_{i,j}})+M(\overline{\mu_{win,j}}\cup\overline{\mu_{i,j}})} \tag{35}$$

where $\cap$ and $\cup$ denote the intersection and union of two fuzzy sets $\tilde{\mu}_{win}, \tilde{\mu}_i$ and the union of two fuzzy sets can be formed as $M(\underline{\mu_{win,j}}\cup\underline{\mu_{i,j}})=M(\underline{\mu_{win,j}})+M(\underline{\mu_{i,j}})-M(\underline{\mu_{win,j}}\cap\underline{\mu_{i,j}})$. $M(.)$ stands for the size of the fuzzy set. This scheme is also committed to $M(\overline{\mu_{win}}\cup\overline{\mu_i})$.Because of the highly nonlinear contour of the Gaussian function, the computation of the union and intersection of two Gaussian fuzzy sets is highly complex. Hence, we approximate it using a triangular membership function, as done in [33],[34]. In what follows, we can thus specify $M(\underline{\mu_{win,j}})=\int_{-\infty}^{\infty}\exp(-\frac{(x-c)^2}{2\sigma^2})dx=\sigma_{win,j}{}^1\sqrt{2\pi}$. As a result of the alignment procedure $c_{win}{}^1=c_i{}^1$, we quantify $M(\underline{\mu_{win,j}}\cap\underline{\mu_{i,j}})$ as follows:

$$M(\underline{\mu_{win,j}}\cap\underline{\mu_{i,j}})=\frac{h^2}{2}+\frac{h^2((\sigma_{win,j}{}^1-\sigma_{i,j}{}^1))}{2(\sigma_{i,j}{}^1-\sigma_{win,j}{}^1)}-\frac{h^2(\sigma_{win,j}{}^1+\sigma_{i,j}{}^1)}{2(\sigma_{win,j}{}^1-\sigma_{i,j}{}^1)} \tag{36}$$

where $h=\max[0,x]$.Performing similar mathematical operations for $M(\overline{\mu_{win,j}}\cap\overline{\mu_{i,j}})$ and $M(\overline{\mu_{win,j}}\cup\overline{\mu_{i,j}})$, we can eventually arrive at $s_{1,j}(win,i)\in[0,1]$ in (31).

Conversely, we apply an extended kernel-based metric approach, which was proposed for the type-1 fuzzy set in [35], to enumerate the distance-based similarity measure $s_{2,j}(win,i)\in[0,1]$. As with $s_{1,j}(win,i)\in[0,1]$, the average cardinality concept is implemented to infer the resultant distance-based similarity measure as follows:

$$S_{2,j}(win,i)=\frac{\exp(-A)+\exp(-B)}{2} \tag{37}$$

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

The kernel-based metric method features the following appealing properties.

$$S_{2,j}(A,B)=1\Leftrightarrow|C_A-C_B|+|\sigma_A-\sigma_B|=0\Leftrightarrow C_A=C_B\wedge\sigma_B=\sigma_B \tag{38}$$

$$S_{2,j}(A,B)<\varepsilon\Leftrightarrow|C_A-C_B|>\delta\vee|\sigma_A-\sigma_B|>\delta \tag{39}$$

Henceforth, two fuzzy rules are deemed identical, if their similarity in the rule level surpasses a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

predefined constant ρ_3. The similarity in the rule level can be elicited by executing $s_{v,j}(win,i)$ for each input dimension and then by combining them with the t-norm operator as follows:

$$S_v \geq \rho_3, S_v = \min_{j=1,..\mu}(s_{v,j}) \tag{40}$$

where $\rho_3 \in [0,1]$ stands for a predefined constant simply fixed as $\rho_3 = 0.5$. This parameter is not problem-specific and this claim is later confirmed in Section IV.

The blow-up effect [31] can ensue when different orientation or non-homogeneous clusters are merged. The merging of non-homogeneous clusters is not recommended, because it possibly inflicts an over-sized volume of the merged cluster. In light of this issue, the volume of the merged cluster should be examined to capture non-homogenous clusters and the merging process is withheld when the volume of the merged cluster exceeds the total volume of two independent clusters as a precursor of the cluster delamination. This situation is defined as follows:

$$V_{merged}{}^1 + V_{merged}{}^2 \leq u((V_{win}{}^1 + V_i{}^1) + (V_{win}{}^2 + V_i{}^2)) \tag{41}$$

We consolidate the input dimension *u* in this case to hedge the curse of dimensionality problem. The fuzzy rules are merged, if both (40) and (41) are satisfied in the training process. We exploit the weighted average strategy to merge two redundant clusters, since it is thought that a more dominant rule, possessing more supports, should be more influential to form the final shape of the merged cluster.

$$C_{merged}{}^{1,2} = \frac{C_{win}{}^{1,2} N_{win}{}^{old} + C_i{}^{1,2} N_i{}^{old}}{N_{win}{}^{old} + N_i{}^{old}} \tag{42}$$

$$\Sigma^{-1}{}_{merged}{}^{1,2} = \frac{\Sigma^{-1}{}_{win}{}^{1,2} N_{win}{}^{old} + \Sigma^{-1}{}_i{}^{1,2} N_i{}^{old}}{N_{win}{}^{old} + N_i{}^{old}} \tag{43}$$

$$N_{merged}{}^{new} = N_{win}{}^{old} + N_i{}^{old} \tag{44}$$

It can be seen that each rule is weighted by its population, thus signifying greater impact of a more populated cluster on the orientation and shape of the merged cluster. On the other hand, the merging of the rule consequent is inspired from the notion of Yager's participatory learning [36].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

It is merely switched on if the rule premise and the rule consequent are contradictory. Two fuzzy rules are deemed contradictory provided that their rule premises are similar, but possess the distinct local sub-models. To this end, the similarity between two local sub-models ought to be first elicited to delve the contradictory degree, where it can be undertaken by measuring the angle created by the two rule consequents. Noticeably, the first order Chebyshev polynomial portrays the trend of the rule consequent in the output space, whereas the higher order constituents form non-linear oscillations. Therefore, we merely focus on the first order component in measuring the angle. The similarity measure of the rule output is formulated as follows:

$$\hat{\phi}_{l,r} = \max_{o=1,\ldots,m}(\phi_{o,l,r}) \;\; \phi_{o,l,r} = \arccos\left(\left|\frac{a_{o,l,r}{}^{T} b_{o,l,r}}{\left|a_{o,l,r}\right\| b_{o,l,r}\right|}\right) \tag{45}$$

$$S_{out}(\Omega_i{}^{l,r}, \Omega_{i+1}{}^{l,r}) = \begin{cases} 1-\frac{2}{\pi}\hat{\phi}_{l,r}, \hat{\phi}_{l,r} \in \left[0, \frac{\pi}{2}\right] \\ \frac{2}{\pi}(\hat{\phi}_{l,r} - \frac{\pi}{2}), \hat{\phi}_{l,r} \in \left[\frac{\pi}{2}, \pi\right] \end{cases} \tag{46}$$

where $a, b \in \Re^{m \times u}$, $a = [w^{o,l,r}{}_{win,1}, w^{o,l,r}{}_{win,3}, \ldots, w^{o,l,r}{}_{win,2u-1}]$ and $b = [w^{o,l,r}{}_{i,1}, w^{o,l,r}{}_{i,3}, \ldots, w^{o,l,r}{}_{i2u-1}]$. Note that we apply the maximum operator to resolve the similarity measure of the rule consequents in the multi-category problems. The Yager's participatory learning-like consequent merging is thus defined as follows:

$$\Omega_{merged}{}^{l,r} = \Omega_{win}{}^{l,r} + \gamma\delta(\Omega_{win}{}^{l,r} - \Omega_i{}^{l,r}), \gamma = \frac{N_{win}{}^{old}}{N_{win}{}^{old} + N_i{}^{old}} \tag{47}$$

$$\delta = \begin{cases} 1, S_v \le S_{out} \\ 0, S_v > S_{out} \end{cases} \tag{48}$$

where $\gamma \in [0,1]$ can be assumed as the basic learning rate and δ can be supposed to be the compatibility measure between the model, whereas the arousal index is kept constant as 0 and $N_{win} > N_i$. It is worth stressing that we solely check the similarity of winning rule, because the rule premise adaptation provided to the winning rule is the major reason of rule redundancy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

It is worth noting notwithstanding being scattered properly during the initialization process, the winning cluster can significantly overlap another rule notably when the next training samples fill up the gap between two clusters. The rule merging scenario should thus take place apart from the conventional rule pruning scenarios as exhibited by T2P+ and T2ERS methods. As with the rule pruning strategy, the rule merging strategy is carried out, if a new rule is not grown - (13), (16) are not satisfied. The underlying rationale is that of the rule premise adaptation (25)-(27), which may cause the winning rule to be overlapping to other rules. This strategy is useful to reduce the computational complexity, because a low risk redundancy is observed when adding new fuzzy rules with the use of fuzzy rule initialization strategy as outlined in Section III.2.

Fuzzily Weighted Generalized Recursive Least Square (FWGLRS) method: the FWGRLS was proposed in our past work [24], where it delineates a local learning version of the GRLS method in [37]. However, this method has not been applied to adjust the interval type-2 fuzzy rule. The salient trait of the FWGRLS method over the standard RLS method lies on an explicit weight decay term, forcing the weight vector to hover around a small bounded range, thus improving the generalization and compactness of the rule base. Since the weight vector is supposed to revolve around the small range, this technique assists the T2ERS method (27) in detecting the inconsequential fuzzy rule. Another paramount property of the FWGRLS method in comparison with other type-2 rule learning methods in [13]-[19] is the local learning concept, where each rule consequent is separately adapted and is assigned with a unique output covariance matrix, thereby leading to more stable adaptation. Accordingly, the learning mechanisms of a particular rule do not affect the convergence and stability of other rules. This method is written as follows:

$$\psi^{l,r}(n) = \Psi_i^{l,r}(n-1)F(n)\left(\frac{\Delta(n)}{\tilde{\Lambda}_i(n)} + F(n)\Psi_i^{l,r}(n-1)F^T(n)\right)^{-1} \tag{49}$$

$$\Psi_i^{l,r}(n) = \Psi_i^{l,r}(n-1) - \psi^{l,r}(n)F(n)\Psi_i^{l,r}(n-1) \tag{50}$$

$$\Omega_i^{l,r}(n) = \Omega_i^{l,r}(n-1) - \varpi\Psi_i^{l,r}(n)\nabla\xi(\Omega_i^{l,r}(n-1)) + \Psi_i^{l,r}(n)(t(n) - y(n)) \tag{51}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

$$y(n) = x_{en}\Omega_i^{l,r}(n) \text{ and } F(n) = \frac{\partial y(n)}{\partial \Omega_i^{l,r}(n)} = x_{en} \tag{52}$$

where $\tilde{\Lambda}_i(n) \in \Re^{(P+1)\times(P\times 1)}$ denotes a diagonal matrix, whose diagonal elements comprise the firing strength of fuzzy rule $\tilde{R}_i$; the covariance matrix of the modeling error is labeled $\Delta(n)$, which is set as an identity matrix [37]. Furthermore, $\nabla\xi(\Omega_i^{l,r}(n-1))$ illustrates the gradient of the weight decay function, which can be determined as any nonlinear function, and may have an inexact gradient solution. Hence, it is extended to the *n-1* time step, whenever the gradient solution is too difficult to be obtained. We choose the quadratic weight decay function $\xi(y_i(n-1)) = \frac{1}{2}(\Omega_i(n-1))^2$ in this paper, because it is capable of shrinking the weight vector proportionally to its current value, whereas $\varpi \approx 10^{-15}$ indicates a predefined constant.

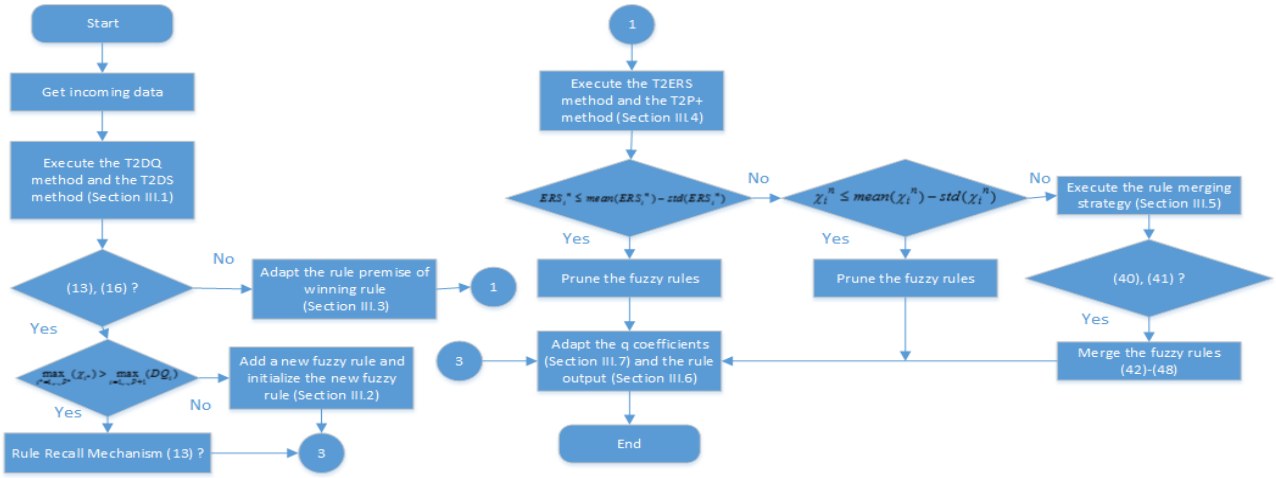

Fig.3 Flowchart of eT2Class

Table 1. Sensitivity of Predefined Parameters

ρ_3	*EC*	*I=1(0.1)*	*I=2(0.3)*	*I=3(0.5)*	*I=4(0.7)*	*I=5(0.8)*	*I=6(0.9)*	sensitivity
iris	CR	0.95±0.3	0.95±0.3	0.95±0.3	0.96±0.2	0.96±0.2	0.97±0.2	0.12
	R	2	2	2	2.1±0.3	2.1±0.3	2.5±0.1	0.13
	RT	0.2±0.05	0.2±0.05	0.2±0.05	0.3±0.05	0.3±0.05	0.35±0.02	0.17
Line	CR	0.94±0.1	0.94±0.1	0.94±0.1	0.94±0.1	0.94±0.1	0.96±0.1	0.17
	R	1.5±0.5	1.5±0.5	1.5±0.5	1.5±0.5	1.5±0.5	2.1±0.5	0.17
	RT	0.27±0.08	0.27±0.08	0.27±0.08	0.27±0.08	0.27±0.08	0.28±0.08	0.17
circle	CR	0.9±0.05	0.9±0.05	0.9±0.05	0.9±0.05	0.9±0.05	0.91±0.03	0.17
	R	1.2±0.4	1.2±0.4	1.2±0.4	1.2±0.4	1.2±0.4	1.3±0.3	0.17
	RT	0.18±0.07	0.18±0.07	0.18±0.07	0.18±0.07	0.18±0.07	0.19±0.04	0.17

CR= Classification rate, R=Rule, RT=Runtime

Adaptation of q Coefficients: the design coefficients $[q_l, q_r]$ are used to control the proportion of upper and lower outputs and should be updated to reflect the uncertainty of the system being classified

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

[19]. In contrast to [17], [18], our proposed method is derived from the Zero-Error Density Maximization (Z-EDM) principle [38], which can produce more reliable predictions for high order statistical behavior than the standard Mean Square Error (MSE) concept. The minimization of the error entropy aims to minimize the distance between the probability distribution of the true class labels and the classifier outputs, thus implicitly inducing the system error to converge to zero. Since the model of the error entropy distribution is unknown, the cost function can be formed using the Parzen window estimation method as follows:

$$\hat{f}(0) = \frac{1}{Nh\sqrt{2\pi}} \sum_{n=1}^{N} \exp(-\frac{{e_{n,o}}^2}{2h^2}) = \frac{1}{Nh\sqrt{2\pi}} \sum_{n=1}^{N} K(\frac{-{e_{n,o}}^2}{2h^2}) \tag{53}$$

where N stands for the number of training observations and h labels the smoothing parameter, which is assigned as 1, and where $e_{n,o}$ presents the system error in the *n-th* training cycle of the *o-th* class. The optimization procedure itself is done with the gradient ascent procedure, which is mathematically defined as follows:

$${q_{i,r}}^o(N) = {q_{l,r}}^o(N-1) + \eta_o \frac{\partial \hat{f}(0)}{\partial {q_{l,r}}^o} = {q_{l,r}}^o(N-1) - \eta_o \frac{1}{N\sqrt{2\pi}} \sum_{n=1}^{N} K(-\frac{{e_{n,o}}^2}{2}) \frac{\partial E}{\partial {q_{l,r}}^o} \tag{54}$$

where η_o is the adaptive learning rate, set using the Lyapunov stability concept. Since the eT2Class runs on the sequential learning environment, (54) should be amended as follows:

$$\sum_{n=1}^{N} \exp(\frac{-{e_n}^2}{2}) = A_N = A_{N-1} + \exp(\frac{-{e_{N,o}}^2}{2}), \frac{\partial \hat{f}(0)}{\partial {q_{l,r}}^o} = \frac{A_N}{N\sqrt{2\pi}} \frac{\partial E}{\partial {q_{l,r}}^o} \tag{55}$$

The gradient term $\partial E / \partial {q_{l,r}}^o$ can be derived by the chain rule as follows:

$$\frac{\partial E}{\partial q_l} = e_{N,o} \frac{\sum_{i=1}^{P} x_e \Omega_{i,o}{}^l (R_i - \overline{R_i})}{\sum_{i=1}^{P} (R_i - \overline{R_i})}, \frac{\partial E}{\partial q_r} = e_{N,o} \frac{\sum_{i=1}^{P} x_e \Omega_{i,o}{}^r (-R_i + \overline{R_i})}{\sum_{i=1}^{P} (-R_i + \overline{R_i})} \tag{56}$$

The learning rate η_o should be tuned in such an avenue to expedite the convergence. To this end, the learning rate is set as follows:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

$$\eta_o(N)=\begin{cases}\rho_5\eta_o(N-1), & \hat{f}(0)^N \geq \hat{f}(0)^{N-1}\\ \rho_4\eta_o(N-1), & \hat{f}(0)^N < \hat{f}(0)^{N-1}\end{cases} \quad \text{,where } 0<\rho_4<1<\rho_5 \tag{57}$$

where $\rho_5 \in (1,1.5]$, $\rho_4 \in [0.5,1)$ indicate the learning rate factors, which steer the direction of the learning rate. Because they are not problem-specific as confirmed in [39], we can assign $\rho_4 = 1.1$ and $\rho_5 = 0.9$. This setting follows the recommendation of [38], where the learning rate should grow, when current cost function at the same time increases and vice versa. Consequently, we should analyze a stable range of η_o to guarantee the convergence.

Theorem 1: suppose that η_o expresses the learning rate of the design factors $[q_l, q_r]$ and $P_{o,\max}$ is defined as the maximum gradient of the output to the design factors $P_{o,\max} = \max_{n=1,..,N} \frac{\partial \hat{y}(n)^o}{\partial {q_{l,r}}^o}$. The asymptotic convergence can be attained by determining the learning rate as $0<\eta_o<\frac{2N\sqrt{2\pi}}{(P_{o,\max})^2 A_N}$.

Proof: the Lyapunov function is defined as $V(k)=e^2(k)/2$, thus landing on the rate o f the Lyapunov function.

$$\Delta V(k)=V(k+1)-V(k)=\frac{1}{2}(e^2(k+1)-e^2(k))=\frac{1}{2}(e(k+1)+e(k))(e(k+1)-e(k)) \tag{58}$$

$$=\frac{1}{2}(e(k+1)+e(k))\Delta e(k)=(e(k)+\frac{1}{2}\Delta e(k))\Delta e(k) \tag{59}$$

The rate of the system error can be derived as follows:

$$\Delta e(k)=e(k+1)-e(k)=\frac{\partial e(k)}{\partial {q_{l,r}}^o}^T \Delta {q_{l,r}}^o \tag{60}$$

$$\Delta {\gamma_i}^o=-\eta_o \frac{A_N}{N\sqrt{2\pi}}\frac{\partial E}{\partial {q_{l,r}}^o}=-\eta_o \frac{A_N}{N\sqrt{2\pi}} e(n)^o \frac{\partial \hat{y}^o}{\partial {q_{l,r}}^o} \tag{61}$$

Therefore, the rate of the Lyapunov function can be established as follows:

$$\Delta V(k)=-\frac{1}{2}\left\|P_o\right\|^2 \eta_o \frac{A_N}{N\sqrt{2\pi}} e_o(n)^2 (2-\left\|P_o\right\|^2 \frac{A_N}{N\sqrt{2\pi}}\eta_o)=-e_o(n)^2 \vartheta \tag{62}$$

Seemingly, the asymptotic convergence can be achieved by $0<\vartheta$, which confirms $0<\eta_o<\frac{2N\sqrt{2\pi}}{(P_{o,\max})^2 A_N}$ and completes the proof at once. The working principle of eT2Class is visualized by a flowchart in Fig.3.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

IV. Sensitivity Analysis of Predefined Parameters

The predefined parameters of eT2Class are argued as a technical flaw, because it may entail expert knowledge or a trial and error process to determine their suitable values. In this section, the sensitivities of these parameters are investigated to justify our claim that these parameters are not sensitive to the learning performance. Nonetheless, we only probe the sensitivity of ρ_3, since other predefined parameters have been well-studied in our past work [39]. To this end, we benefit from the sensitivity analysis of [40], which is illustrated as follows:

$$sensitivity = \frac{1}{k(\pi_{\max} - \pi_{\min})} \sum_{j=1}^{k-1} (EC({\rho_3}^{j+1}) - EC({\rho_3}^{j})) \tag{63}$$

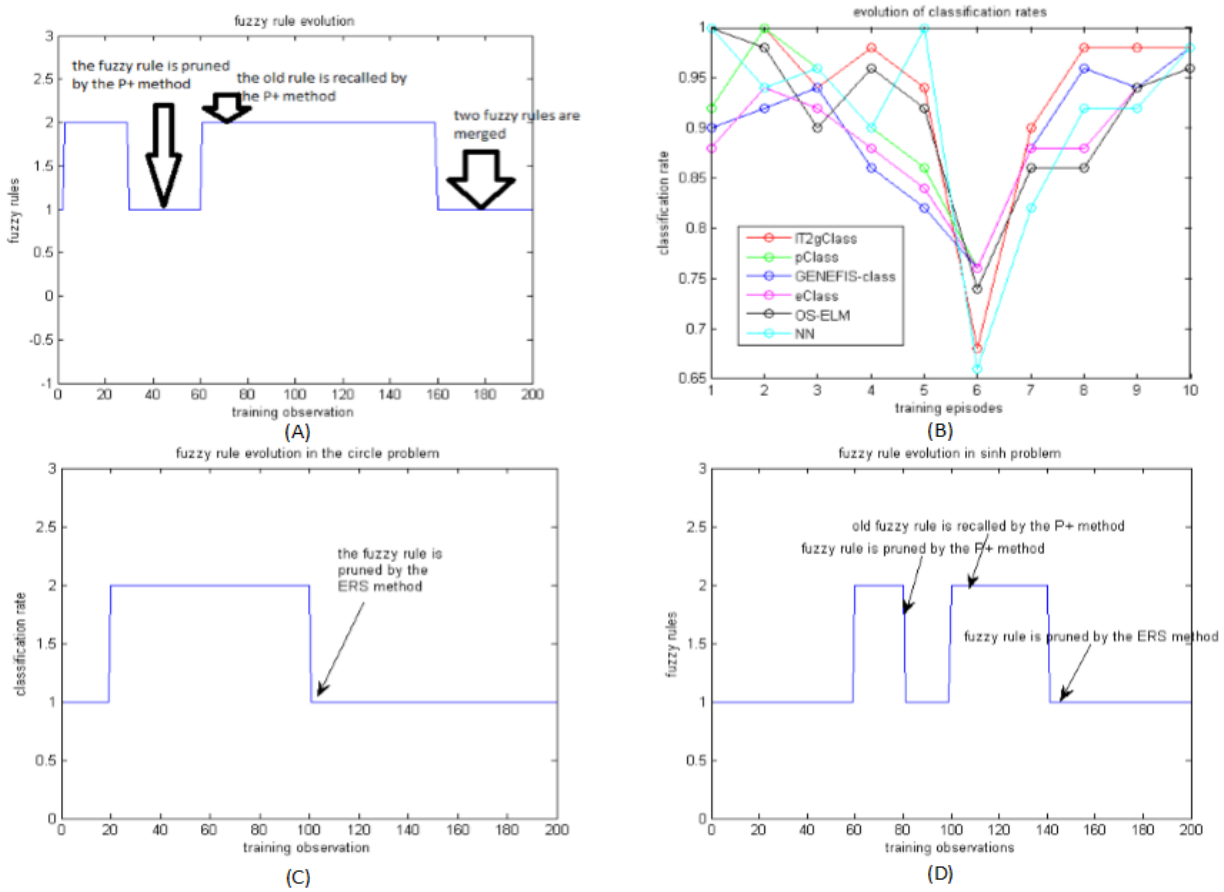

Fig.4 (a) fuzzy rule evolution in the line dataset, (b) the evolution of Classification rates in the line dataset, (c) fuzzy rule evolution in the circle dataset, (d) fuzzy rule evolution in the sinh dataset

where *k* stands for the number of variations or experiments used to test ρ_3, which is stipulated as six, whereas *EC* denotes the evaluation criteria. Meanwhile, $\pi_{win}, \pi_{\max}$ shows the maximum and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

minimum evaluation criteria across six experiments. Because we utilize three evaluation criteria: the classification rates; the number of fuzzy rules; and the runtime, $\pi_{win}, \pi_{\max}$ refer to the maximum and minimum numerical results of each criterion over six trials. We vary $\rho_3 = [0.1, 0.3, 0.5, 0.7, 0.8, 0.9]$ to study their virtual effects on the eT2Class numerical results. Our analysis is conducted by using three data streams: the iris data stream, from the UCI machine learning repository (http://www.ics.uci.edu/mlearn/MLRepository.html); the line and circle data streams from the Diversity in Dealing with Drift (DDD) data base. Our experiments are carried out by making use of the tenfold cross validation technique [41].

Obviously, the predefined threshold ρ_3 is not problem-specific in which different values of this parameter do not substantially affect the performance of eT2Class. For simplicity, we set $\rho_3 = 0.5$ for all our simulations in this paper.

V. Numerical Examples

The efficacy of the eT2Class is numerically validated by committing rigorous empirical studies in real-world and artificial problems, characterizing various types of concept drift. The synthetic problems are paramount to analyze the performance of classifiers, because we can observe the type of drift and the precise time period when the concept drift is active. Nine study cases, namely SEA from [42], electricity pricing, hyper-plane from [48], and 10dplane, sinh, line circle, boolean from DDD database [43], are exploited to assess the learning performance of the eT2Class. All of which characterize various concept drifts, which are very relevant to test the efficacy of EFC. In addition, our own synthetic data stream in [8] is included to validate our algorithm, because it contains noisy and non-stationary properties and dynamic class labels, which are difficult to be handled by EFC.

The efficacy of the eT2Class is benchmarked with five state-of-the-art classifiers: Evolving Classifier (eClass) [4], Parsimonious Classifier (pClass) [8], Generic Evolving Neuro-Fuzzy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

Inference System Classifier (GENEFIS-class) [7], Online Self Organizing Extreme Learning Machine (OS-ELM) [44], and Neural Networks [3]. eClass, pClass, and GENEFIS-class are subsumed in the class of evolving classifiers, adopting an open structure concept as with eT2Class. OS-ELM exhibits an incremental learning scenario with the absence of the automatic knowledge building mechanism. On the other hand, NN characterizes a classical batched algorithm, which is iterative in nature. It has to revisit the preceding training data and retrain itself according to a specified number of epochs to generate an optimal decision boundary.

Table 2. The numerical results of consolidated classifiers

ALGORITHMS		eT2Class	pClass	eClass	GENEFIS-class	OS-ELM	NN
SEA dataset	Classification rate	**0.81±0.2**	0.78±0.04	0.76±0.03	0.76±0.01	0.61±0.001	0.86±0.03
	Rule	**1.3±0.5**	3.5±2.42	15.9±3.4	2.9±1	100	100
	Time(s)	4.4±1.4	3.86±0.5	10.07±2.3	3.02±0.26	**0.006±0.008**	3.27±0.6
	Rule base	**37.7**	70	190.8	58	600	600
Electricity pricing dataset	Classification rate	**0.77±0.1**	0.7±0.1	0.76±0.07	0.75±0.0	0.57±0.09	0.53±0.08
	Rule	**2.3±0.5**	3.5±2.4	11.9±0.07	3.5±1.5	100	100
	Time(s)	6.6±0.8	3.97±0.91	4.12±2.2	**0.49±0.4**	2.43±0.2	9.61±2.2
	Rule base	354.2	917	321.3	315	1100	1100
10dplane dataset	Classification rate	**0.83±0.2**	0.68±0.02	0.68±0.2	0.68±0.01	0.75±0.05	0.68±0.18
	Rule	**2.3±0.5**	5.6±0.63	10.1±3.6	5.6±0.12	100	100
	Time(s)	0.2±0.004	0.17±0.08	0.12±0.09	0.14±0.05	0.09±0.02	0.66±0.63
	Rule base	687.7	873.6	**363.3**	873.6	1400	1400
Circle dataset	Classification rate	**0.91±0.04**	0.73±0.17	0.7±0.11	0.7±0.03	0.66±0.14	0.69±0.02
	Rule	**1.2±0.42**	2.8±1.1	3.6±0.84	3.2±1.03	50	800
	Time(s)	0.18±0.07	0.15±0.008	0.09±0.01	0.15±0.01	0.08±0.02	6.7±0.8
	Rule base	**24**	33.6	**32.4**	38.4	500	5600
Line dataset	Classification rate	**0.94±0.1**	0.91±0.07	0.89±0.06	0.9±0.07	0.91±0.08	0.89±0.1
	Rule	**1.1±0.3**	2.5±0.71	4.4±0.51	3.6±0.7	25	200
	Time(s)	0.13±0.04	0.15±0.0009	0.1±0.009	0.14±0.01	**0.04±0.02**	0.82±0.12
	Rule base	**22**	30	39.6	43.2	250	1400
Sinh dataset	Classification rate	**0.71±0.03**	0.71±0.09	0.7±0.07	0.71±0.06	0.68±0.04	0.69±0.12
	Rule	**1.3±0.5**	3.6±1.9	6.3±1.5	3.6+0.8	50	500
	Time(s)	0.2±0.07	0.17±0.01	0.13±0.02	0.15±0.02	**0.07±0.02**	0.6±0.02
	Rule base	**26**	43.2	56.7	43.2	500	3500
Weather dataset	Classification rate	**0.81±0.01**	**0.8±0.01**	0.8±0.05	0.8±0.02	0.74±0.06	0.8±0.07
	Rule	**2.3±0.8**	**3.8±2.5**	5.6±1.72	4.4±1.64	80	80
	Time(s)	1.7±0.2	1.27±0.18	1.13±0.3	1.13±0.14	**0.56±0.7**	1.4±0.1
	Rule base	354.2	342	**151.2**	396	2160	2160
Hyper-plane dataset	Classification rate	**0.93±0.02**	0.92±0.02	0.91±0.02	0.91±0.01	0.88±0.03	0.91±0.09
	Rule	**1.2±0.4**	2.2±0.63	8.6±2	3.39±0.12	35.3±4.16	10
	Time(s)	2.1±0.4	1.86±0.07	13.48±3.61	3.4±0.05	**1.22±0.13**	5.65±0.06
	Rule base	**55.2**	66	124.4	90	2118	70
Noise corrupted signal dataset	Classification rate	**0.75±0.1**	0.74±0.12	0.72±0.12	0.73±0.09	0.72±0.14	0.68±0.3
	Rule	**1.3±0.5**	3±1.2	3.7±1.3	4.5±1.1	50	80
	Time(s)	9.1±3	6.4±0.7	6.9±1.9	7.5±0.9	**2.23±0.11**	10.6±0.3
	Rule base	**11.7**	24	29.6	36	400	640
Boolean dataset	Classification rate	**0.91±0.16**	0.83±0.2	0.81±0.2	0.82±0.2	0.8±0.17	0.8±0.1
	Rule	**1.3±0.5**	2.6±0.8	3.7±1.1	2.6±1.1	100	100
	Time (s)	0.09±0.03	0.08±0.002	**0.04±0.01**	0.09±0.05	0.24±0.02	0.2±0.02
	Rule base	**20**	52	44.4±12.7	52	600	600

Three types of data streams are available for each problem in the DDD database. In our numerical studies, we employ the most complex version, where the most samples suffer from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

changing data distributions. The consolidated classifiers are evaluated in four viewpoints: the classification rate of the testing phase; the number of fuzzy rules generated in the training process; the execution time to accomplish the training process; and the number of network parameters to be saved in the memory. The network parameters are computed as follows: $O(p\times m\times(2u+1)+p\times(u\times u)+p\times u)$ (gClass), $O(UP+P+mP(U+1)+P)$ (eClass), $O(p\times m\times(2u+1)+p\times(u\times u)+p\times u)$ (pClass, GENEFIS-class), and the eT2Class $O(p\times m\times(2u+1)+2p\times(u\times u)+p\times u)$. We rely on an Intel (R) core (TM) i7-2600 CPU @3.4 GHz processor and 8 GB memory to carry out simulations. The predefined parameters of the benchmarked classifiers are set according to the rule of thumb in their original publications. The periodic hold-out procedure [8] simulating the training and testing processes in real-time is exploited as the experimental procedure. Therefore, Table 2 encapsulates the average numerical results of the benchmarked classifiers across the periodic hold-out procedure. Fig.4(a) visualizes the fuzzy rule evolution of the eT2Class in the line dataset, while Fig.4(b) pictorially displays the trace of the classification rates of all classifiers in the line dataset. Fig.4(c) and (d) depict the trace of fuzzy rules in the circle and sinh problems.

Table 3. Ranking of Classifiers.

Study cases	eT2Class (CR, R, ET, RB)	pClass (CR, R, ET, RB)	eClass (CR, R, ET, RB)	GENEFIS-class (CR, R, ET, RB)	OS-ELM (CR, R, ET, RB)	NN (CR, R, ET, RB)
Sea dataset	(1,1,5,1)	(2,2,2,2)	(4,4,6,4)	(3,3,3,3)	(6,5,1,5)	(5,5,4,5)
Electricity pricing dataset	(1,1,5,3)	(2,2,2,4)	(3,4,3,2)	(4,3,1,1)	(5,5,4,5)	(6,5,6,5)
10dplane dataset	(1,1,5,2)	(3,3,4,3)	(4,4,2,1)	(2,2,3,4)	(6,5,1,5)	(5,5,6,5)
Circle dataset	(1,1,5,1)	(2,2,3,3)	(4,4,2,2)	(3,3,4,4)	(6,5,1,5)	(5,6,6,6)
Line dataset	(1,1,3,1)	(2,2,5,2)	(5,4,2,3)	(4,3,4,4)	(3,5,1,5)	(6,6,6,6)
Sinh dataset	(1,1,5,1)	(3,3,4,2)	(4,4,2,4)	(2,2,3,3)	(6,5,1,5)	(5,6,6,6)
Weather dataset	(1,1,6,3)	(2,2,4,2)	(4,4,3,1)	(3,3,2,4)	(6,5,1,5)	(5,5,5,5)
Hyper-plane dataset	(1,1,3,1)	(2,2,2,2)	(4,4,6,5)	(3,3,4,3)	(6,6,1,6)	(5,5,5,4)
Noise corrupted signal dataset	(1,1,5,1)	(2,2,2,2)	(4,3,3,3)	(3,4,4,4)	(5,5,1,5)	(6,6,6,6)
Boolean dataset	(1,1,3,1)	(2,2,2,3)	(4,4,1,2)	(3,3,4,4)	(6,5,6,5)	(5,5,5,5)
Average	(**1**,**1**,4.5,**1.5**)	(2.2,2.2,3,2.5)	(4,3.9,3,2.7)	(3,2.9,2.8,3.4)	(5.5,5.1,**1.8**,5.1)	(5.3,5.4,5.5,5.3)

CR=classification rate, R=rule, ET=execution time, RB=rule base

It can be seen from Table 2 that the eT2Class delivers the most reliable classification rates, while retaining the most economical number of fuzzy rules and rule base parameters. Although the interval type-2 fuzzy rule of the eT2Class benefits from the multivariable Gaussian function

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

and the Chebyshev polynomial, which are presumed to incur a costlier memory demand than the conventional interval type-2 fuzzy rules, the eT2Class is capable of landing on the most compact and parsimonious rule bases in 10 study cases. This fact confirms the efficacy of the non-axis parallel ellipsoidal cluster evolved by the generalized interval type-2 fuzzy rule, being able to alleviate the required number of fuzzy rules. The self-organizing property of the eT2Class is visualized by Fig.4(a),(c),(d) where the fuzzy rules can be added, pruned, merged, recalled on demands from data streams. Conversely, the generalization ability of the eT2Class is depicted in Fig.4(b), where the eT2Class produces more encouraging trend of classification rate than other classifiers.

VI. Statistical Test

Our numerical results are further substantiated by thorough statistical tests to draw a statistically valid conclusion about the eT2Class learning performance. Table 3 summarizes the classifier rankings, in which the eT2Class overcomes the other classifiers on three criteria-the classification rate, the number of fuzzy rules, the number of rule base parameters, whereas the eT2Class is inferior to those of eClass, pClass, GENEFIS-class, and OS-ELM in the aspect of the execution times.

Table 4. Difference in performance between the eT2Class and other algorithms

Algorithms	CR	R	RB	ET
eT2Class vs eClass	3.59	3.47	1.43	1.79
eT2Class vs GENEFIS	2.39	2.27	2.27	2
eT2Class vs pClass	1.43	1.43	1.2	1.79
eT2Class vs OS-ELM	6.57	4.9	4.3	3.22
eT2Class vs NN	5.14	5.24	4.54	-1.19

CR=classification rate, R=rule, ET=execution time, RB=rule base

A Friedman statistical test [45] with $\alpha = 0.1$, 5 degree of freedom and the critical value of 9.23 is first carried out. We arrive at $\chi_F^2 = 44.8,31.8,17.37,22.7$, thus rejecting the null hypothesis outright for all evaluation criteria. Since the Friedman test is deemed conservative, the ANOVA test in [45] as an extension of the Friedman test is executed to bear out the conclusion. Accordingly, the critical value of $\alpha = 0.05$ with (5,45) DOF is 2.42, whereas we elicit $F_F = 77.53,15.72,4.76,7.48$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

Hence, the similar conclusion as elicited in the Friedman statistical test can be achieved, where the null hypothesis is rejected. The main goal of these two statistical tests is to reveal the performance difference among benchmarked classifiers, where the rejection of null hypothesis implies that the performance difference among benchmarked classifiers is noticeable. Nonetheless, these two statistical tests do not disclose that the eT2Class statistically surmounts its counterparts.

The aforementioned issue is unraveled with the use of a post-hoc test of Bonferroni-Dunn of [45], where the eT2Class is said to conquer another classifier, whenever its performance difference beats the critical difference $CD = 1.95$ with the critical value $\alpha = 0.1$. The difference in performance of the eT2Class and another classifier for four evaluation criteria is elicited using $z = (R_i - R_j)\sqrt{6Q}\Big/\sqrt{(M+1)M}$ and is then reported in Table 4. *Q=10* is the number of study cases and *M=6* is the number of consolidated classifiers, whereas R_i, R_j stand for the average rankings of *i* and *j* classifiers respectively. In summary, the eT2Class is more superior to eClass, GENEFIS-Class, OS-ELM, and NN in the criteria of classification rate and fuzzy rule, while outperforming GENEFIS-class, OS-ELM and NN in the standpoint of rule base parameter. On the other hand, eT2Class is only inferior to that of OS-ELM in realm of the execution time.

VII. Conceptual Comparisons

This section conceptually discusses the novel facets of the eT2Class in contrast with other prominent Interval Type-2 EFSs such as [13]-[15], [17], and [18]. Roughly speaking, the eT2Class is more acceptable for the online learning scenario than those in [13] and [14], because [13] and [14] are reliant on the KM-type reduction method, which is iterative in nature. A more trustworthy input partitioning strategy is embedded in the eT2Class, whereas [13],[14] make use of a classical rule firing strength to evolve fuzzy rules. In addition, [13],[14] are defective due to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2015.2463732, IEEE Transactions on Fuzzy Systems

the absence of salient learning modules such as rule pruning, merging, and recall strategies. The eT2Class also constitutes a high-order classifier, whereas [15] actualizes a zero-order classifier. Notwithstanding sequential classifiers, [17], [18] are essentially akin to those of [13],[14], which apply the distance-based rule growing technology and suffer from the absence of the rule pruning, recall, merging modules.

Table 5. Characteristics of consolidated classifiers

Classifiers	Fuzzy rule	Rule Growing	Rule Pruning	Rule merging	Rule Adaptation
eT2Class	Generalized type-2 fuzzy rule	T2DQ and T2DS	T2ERS and T2P+	Vector similarity measure	T2FWGRLS, ZEDM
GENEFIS-class	Generalized type-1 fuzzy rule	DQ, DS, GART+	ERS and P+	Kernel-based metric	FWGRLS
pClass	Generalized type-1 fuzzy rule	DQ, DS	ERS and P+	N/A	FWGRLS
rClass	Generalized type-1 fuzzy rule	DQ, DS	ERS and P+	Geometry measure	FWGRLS, ZEDM

We continue our analysis by comparing eT2Class with its predecessors: GENEFIS-Class [7]; pClass [8]; rClass [39]. The consolidated algorithms are assessed in five viewpoints to illustrate the novelty of eT2Class against our previous works: fuzzy rule; rule pruning; rule merging; rule growing; fuzzy rule initialization; parameter learning. The learning properties of the consolidated algorithms are tabulated in Table 4. Obviously, eT2Class advances our past works to suit to the working framework of the type 2 fuzzy system. In realm of fuzzy rule, other classifiers make use of the same fuzzy rules, whereas eT2Class employs the interval type-2 multivariable Gaussian function in the rule premise and the interval type-2 Chebyshev function in the rule consequent. Another important contribution of eT2Class can be found in the rule growing method and the rule pruning scenario, in which DQ, DS, ERS and P+ methods are modified to deal with the type-2 fuzzy rule. The same strategy occurs in the FWGRLS method to accommodate the type-2 fuzzy rule. The ZEDM method of eT2Class is unlike in rClass, because it is used to adapt the design factors - the type reduction mechanism. Furthermore, the rule merging scenario of eT2Class is derived from the vector similarity measure, which is an uncharted territory by any our past works.

VIII. CONCLUSION AND FURTHER STUDY

To address three learning flaws of the currently applied evolving systems, a new evolving interval type-2 fuzzy rule-based classifier, namely Interval Type-2 Generic Classifier (eT2Class), is presented in this paper. The eT2Class realizes a fully evolvable classification paradigm under the interval type-2 fuzzy platform. The viability of the eT2Class is numerically validated by rigorous study cases, and benchmarks, involving state-of-the art classifiers and statistical tests. In summary, the eT2Class delivers the most reliable learning performance in attaining a trade-off between complexity and simplicity. The performance of the eT2Class will be corroborated in the future work by using the meta-cognitive-based scaffolding theory in cognitive psychology. We will also develop an application of the eT2Class in customer churn prediction and management.

ACKNOWLEDGEMENTS

The work presented in this paper is partly supported by the Australian Research Council (ARC) under Discovery Projects DP110103733 and DP140101366 and the first author acknowledges receipt of UTS research seed funding grant.

REFERENCES

[1] 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)

[2] E. Lughofer and P. Angelov, Handling drifts and shifts in on-line data streams with evolving fuzzy systems, *Applied Soft Computing*, vol. 11(2), pp. 2057-2068, (2011)

[3] S. Haykin, *Neural Networks: A Comprehensive Foundation* (2nd edn), Prentice Hall., Upper Saddle River, New Jersey, 1999

[4] P. Angelov and X. Zhou, "Evolving fuzzy-rule-based classifiers from data streams," *IEEE Transactions on Fuzzy Systems*, vol. 16(6), pp. 1462-1475, (2008)

[5] E.Lughofer, O.Buchtala, " Reliable all-pairs evolving fuzzy classifiers", *IEEE Transactions on Fuzzy Systems*, vol. 21(4), pp. 625-641, (2013)

[6] A. Lemos, W. Caminhas and F. Gomide, Adaptive fault detection and diagnosis using an evolving fuzzy classifier, *Information Sciences*, vol. 220, pp. 64--85, (2013)

[7] M.Pratama, S.Anavatti, E.Lughofer, "GENFIS: Towards an effective localist network", *IEEE Transactions on Fuzzy Systems*, on line and in press, (2013)

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

[9]H-J.Rong, N.Sundarajan, G-B.Huang, G-S Zhao, " Extended Sequential Adaptive Fuzzy Inference System for Classification Problems", *Evolving System*, Vol.2 (2), pp.71-82, (2011)

[10] 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.

[11] 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)

[12] 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)

[13] 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)

[14] 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)

[15] 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)

[16] 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)

[17] 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)

[18]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)

[19] 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

[20]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)

[21]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)
[22]D.Wu, J.M.Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets", *Information Sciences*, Vol.179, pp.1169-1192, (2009)
[23] 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)
[24] M.Pratama, S.Anavatti, E.Lughofer, " Evolving fuzzy rule-based classifier based on GENEFIS", *Proceedings of the IEEE Conference on Fuzzy Systems*, Hyderabad, India, 2013
[25]Y.H.Pao, " *Adaptive Pattern Recognition and Neural Networks*", Reading, MA: Addison-Wesley, 1989
[26] J.C.Patra, R.N.Pal, B.N.Chatterji, G.Panda, " Identification of nonlinear dynamic systems using functional link artificial neural networks," *IEEE Transactions on Systems, Man and Cybernetics*, Vol.29(2), pp.254-262, (1999)
[27] 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)
[28] M.Pratama, S.Anavatti, E.Lughofer, C-P.Lim, " gClass: An Incremental Meta-Cognitive-based Scaffolding Theory", *Submitted to a special issue on IEEE Computational Intelligence Magazine* , 25th of August (2014)
[29] K. Tabata, M. S. M. Kudo, Data compression by volume prototypes for streaming data, *Pattern Recognition*, Vol.43(9),pp. 3162—3176, (2010)
[30]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)
[31] E. Lughofer, *Evolving Fuzzy Systems --- Methodologies, Advanced Concepts and Applications*, Springer, Heidelberg, (2011)
[32]S.W.Tung, C.Quek, C.Guan, "eT2FIS: An Evolving Type-2 Neural Fuzzy Inference System", *Information Sciences*, vol.220, pp.124-148, (2013)
[33] C. F. Juang and C. T. Lin, "An on-line self-constructing neural fuzzy inference network and its applications," *IEEE Transactions on Fuzzy Systems.*, vol. 6,no. 1, pp. 12–32, Feb. 1998.
[34] 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)
[35] 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)
[36] R. R. Yager, "A model of participatory learning," *IEEE Transaction on Systems., Man, Cybernetics.*, vol. 20, no. 5, pp. 1229–1234,(1990)
[37] 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)
[38] K.Subramanian, R.Savitha, S.Suresh, " Zero-Error Density Maximization Based Learning Algorithm for a Neuro-Fuzzy Inference System", *in proceeding of IEEE Conference on Fuzzy System (Fuzz-IEEE)*, Hyderabad, India, (2013)
[39] M.Pratama, J.Lu, S.Anavatti," Recurrent Classifier based on An Incremental Meta-Cognitive-based Scaffolding Algorithm", *submitted to IEEE Transactions on Fuzzy Systems*, 22nd of August, (2014)
[40] C. F. Juang and C. D. Hsieh, "A locally recurrent fuzzy neural network with support vector regression for dynamic system modeling," IEEE Trans. Fuzzy Systems, vol.18, no.2, pp.261-273, (2010)
[41] M.Stone, "Cross-Validatory Choice and Assessment of Statistical Predictions", *Journal of Royal Statistic Society*, vol.36, pp. 111-147, (1974)
[42] 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)
[43] L.LMinku, X.Yao, " DDD: A New Ensemble Approach for Dealing with Drifts", *IEEE Transactions on Knowledge and Data Engineering*, Vol.24(4), (2012)
[44] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan " A Fast and Accurate On-line Sequential Learning Algorithm for Feedforward Networks", *IEEE Transactions on Neural Networks*, vol.17(6), pp.1411-1423, (2006)
[45]J.Demsar, "Statistical Comparisons of Classifiers over Multiple Datasets", *Journal of Machine Learning Research*, Vol.7, pp.1-30, (2006)
[46] P. Angelov, E. Lughofer, and X. Zhou, "Evolving fuzzy classifiers using different model architectures," *Fuzzy Sets and Systems*, vol. 159(23), pp. 3160–3182, (2008)
[47] Yang, X. Zhang, G. Lu J. and Ma J, A kernel fuzzy c-means clustering based fuzzy support vector machine algorithm for classification problems with outliers or noises, *IEEE Transactions on Fuzzy Systems* Vol. 19(1), pp.105-115, (2011)
[48] http://moa.cms.waikato.ac.nz/datasets/