On Line Learning Fuzzy Rule-based System Structure from Data Streams

Plamen Angelov, *Senior Member IEEE*, Xiaowei Zhou, *Student Member IEEE*

Abstract —**A new approach to fuzzy rule-based systems structure identification in on-line (possibly real-time) mode is described in this paper. It expands the so called evolving Takagi-Sugeno (eTS) approach by introducing self-learning aspects not only to the number of fuzzy rules and system parameters but also to the number of antecedent part variables (inputs). The approach can be seen as on-line sensitivity analysis or on-line feature extraction (if in a classification application, e.g. in eClass which is the classification version of eTS). This adds to the flexibility and self-learning capabilities of the proposed system. In this paper the mechanism of formation of new fuzzy sets as well as of new fuzzy rules is analyzed from the point of view of on-line (recursive) data density estimation. Fuzzy system structure simplification is also analyzed in on-line context. Utility- and age-based mechanisms to address this problem are proposed. The rule-base structure evolves based on a gradual update driven by; i) information coming from the new data samples; ii) on-line monitoring and analysis of the existing rules in terms of their utility, age, and variables that form them. The theoretical theses are supported by experimental results from a range of real industrial data from chemical, petro-chemical and car industries. The proposed methodology is applicable to a wide range of fault detection, prediction, and control problems when the input or feature channels are too many.**

I. INTRODUCTION

Fuzzy rule-based models, and in particular, Takagi-Sugeno type of fuzzy rule-based models have found numerous applications during the last couple decades [1,2]. During the first decade (late 1980s, early 1990s) they were primarily designed using expert knowledge and considered a full coverage of the data space [3] (usually equally distributed between the fuzzy sets that represent different linguistic terms). Since mid-1990s a different approach became popular as 'data-driven' fuzzy systems design [4]-[6]. The importance of this approach should not be underestimated because it makes possible automation of the process of fuzzy systems design (including on-line and in real-time). This makes fuzzy systems much more attractive for real applications including in areas such as robotics, advanced industrial processes etc. [7]-[8]. Still, the problems of structure identification of the fuzzy rule based system; especially in on-line mode is paid little or no attention [3,4]. Usually, it is assumed that the structure is selected and pre-fixed. In some works it is a result of data space partitioning by clustering [5], but in off-line mode and with a pre-defined number of clusters and pre-defined number of inputs (features). In the present paper, we challenge this problem by a proposal for a fully data-driven method that allows a truly flexible and evolving structure of the fuzzy systems (in particular of Takagi-Sugeno type) to be designed on-line. The basic principle is of a *gradual evolution* of the fuzzy rule-based model structure in terms of fuzzy rules and their components – fuzzy sets, including variables of the antecedent part. The structure of the fuzzy rules is determined by on-line incremental clustering of the input-output data space. The inputs (features) are gradually selected from the initial pool of potential inputs (features) based on their accumulated sensitivity. The proposed approach is verified on a number of real industrial examples. The quality of the fuzzy rule base can be monitored on-line and simplified based on the removal of fuzzy rules that are old or marginally used.

The remainder of the paper is structured as follows. In Section II the approach to on-line learning the antecedent part of the fuzzy systems by recursively estimating data density distributions is presented. In Section III the newly proposed approach to select inputs (features) on-line is introduced. In Section IV mechanisms for fuzzy rules reduction based on cluster/rule *age* and *utility* are introduced. Section V describes the application of the proposed techniques to various experimental data, and finally, the Conclusions and Discussion Section closes the paper.

Both authors are with the Intelligent Systems Research Laboratory, Department of Communication Systems, InfoLab21, Lancaster University, Lancaster, LA1 4WA, UK; phone +44 (1524) 510391; fax: +44 (1524) 510489, e-mail: p.angelov@lancaster.ac.uk

II. ON-LINE LEARNING FUZZY SYSTEMS ANTECEDENTS BASED ON DATA DENSITY

The main driver of the proposed approach is the attempt to estimate the data density in a recursive (therefore computationally efficient and suitable for on-line and real-time applications) manner from the data streams and to react on the data density variations by modifying the underlying structure of the model. It is well known that the data density has been estimated off-line by kernels in image processing [9], by Parzen windows [10] in statistical learning. It is also exploited in generalized regression models [11], so called Mountain function [12] and so called potential [13]. In all of these methods the data density is estimated based on Gaussian distribution. In [14] it was proposed to use Cauchy function over the sum of distances between data points, because the Cauchy function same as

978-1-4244-1819-0/08/$25.00©2008 IEEE

the Gaussian:

a) is monotonic;

b) its maximum is unique and of value 1;

c) asymptotically tends to zero when the argument tends to infinity.

These properties, which are necessary for a density function, are satisfied by both functions, because the Cauchy function is, in fact, a first order approximation of the Gaussian:

$$e^{-\sum_{i=1}^{k-1}\frac{\|z(k)-z(i)\|^2}{2\sigma^2}} = \frac{1}{e^{\sum_{i=1}^{k-1}\frac{\|z(k)-z(i)\|^2}{2\sigma^2}}} \approx \frac{1}{1+\sum_{i=1}^{k-1}\frac{\|z(k)-z(i)\|^2}{2\sigma^2}+\ldots} \tag{5}$$

where z denotes the input-output vector; k is the current time instant; σ denotes the spread or zone of influence of the cluster.

The data density (expressed by the value called potential, P) has been used as a criterion to form new clusters in the evolving fuzzy clustering approach, eClustering [14]:

$$P(z(k)) = \frac{1}{1+\frac{1}{k-1}\sum_{i=1}^{k-1}\frac{\|z(k)-z(i)\|^2}{2\sigma^2}} \tag{6}$$

The rationale is that the points with high potential are good candidates for becoming focal points of fuzzy rules. Estimation of the data density in on-line mode is not a trivial task, because by definition density in the data space around the current data point is based on the distance between this point and all other data samples which in an on-line mode will not be kept in the memory. This require a recursive procedure to calculate (6), which was proposed in [14,15]:

$$P(z(k)) = \frac{k-1}{(k-1)(a(k)+1)+b(k)-2c(k)} \tag{7}$$

Values $a(k)$ and $c(k)$ can be calculated from the current frame only:

$$a(k) = \sum_{j=1}^{n+m} z_j^2(k);\ c(k) = \sum_{j=1}^{n+m} z_j(k)d_j(k) \tag{8}$$

where n denotes the dimensionality of the inputs; m denotes dimensionality of the outputs; $d_j(k)$ is calculated recursively as shown below.

The value $b(k)$ is also accumulated during the processing of the frames one by one as given by the following recursive expressions:

$$b(k) = b(k-1) + a(k-1); b(1) = 0 \tag{9}$$

$$d_j(k) = d_j(k-1) + z_j(k-1); d_j(k) = 0 \tag{10}$$

The value of the spread (zone of influence of the clusters), σ is proposed to be updated on-line in a data-driven fashion by learning the data distribution and variance [16]:

$$\sigma_{ij}^2(k) = \alpha\sigma_{ij}^2(k-1) + (1-\alpha)\frac{1}{N_i(k)}\sum_{l=1}^{N_i(k)}(z_l(k)-z_l(k))^2 \tag{11}$$

where α denotes the learning step (recommended value 0.5); $N_i(k)$ denotes the number of data samples that are associated with the i^{th} cluster based on the closeness; the initial value of the spread is usually $\sigma_j(1) = 0.5$.

Note that the focal points and their potentials are kept in the memory while all other previous data samples are discarded from the memory. The potentials of the focal points are, however, updated at each time step (with each new data sample being read) because by definition the potential should represent the data density in respect to all data samples (including the ones that will appear after the moment when a sample is being used as a prototype to form a cluster). Therefore, an update of the potential of focal points is required, which can be done by the following formula [14-16]:

$$P(z_i^*(k)) = \frac{k-1}{k-1+(k-2)\left(\frac{1}{P(z_i^*(k-1))}-1\right)+\left\|x_{ij}^*-x_{ij}\right\|^2} \tag{12}$$

Once the data density is estimated by (7)-(10) with the spread adapted by (11) one can form a fuzzy rule-base according to the following basic principles:

1) a data sample that have <u>high potential</u> is eligible to be a focal point of a fuzzy rule;

2) a data sample that lies in an area of <u>data space **not covered**</u> by other fuzzy rules is also eligible to form a fuzzy rule in order to ensure coverage of the data space;

3) <u>avoid overlap</u> and information redundancy in forming new fuzzy rules.

The first principle, 1) is reflected by the following expression [15,17]:

$$P(z(k)) > \max_{i=1}^{R} P\left(z^*(k)\right) \tag{13}$$

where z^* denotes the data sample that has already been chosen to be a focal point of a fuzzy rule; R denotes the number of fuzzy rules up to the moment k (before condition (13) is checked.

A milder version of (13) can be formulated which makes easier to form new clusters if there are many data points that are not associated to any cluster called 'outer' data points:

$$\gamma P(z(k)) > \max_{i=1}^{R} P\left(z^*(k)\right) \tag{14}$$

Where $\gamma = \begin{cases} \frac{\log k}{N-3} & \mu_j^i(x_k) < e^{-2}; \forall i, \forall j \\ 1 & else \end{cases}$ is a factor that takes into account the 'outer' data points; N is the number of the 'outer' data points.

The second principle, 2) is reflected by the following expression [18]:

$$P(z(k)) < \min_{i=1}^{R} P\left(z^*(k)\right) \tag{15}$$

Based on both principles, 1) and 2) we can formulate two alternative versions of the conditions, namely:

- Condition A1: (13) OR (15);
- Condition A2 (14).

The third principle, 3) and the possibility of the rule-base to gradually shrink is guaranteed by the following **condition B** [16]:

$$\exists i, i = [1, R]; \mu_{ij}\left(x(k)\right) > e^{-1}; \forall j; j = [1, n] \tag{16}$$

where $x = \left[x_1, x_2, ..., x_n\right]^T$ denotes the input vector (features if classification problem); μ_{ij} denotes the membership function (usually of Gaussian type) of the j^{th} fuzzy set of the i^{th} fuzzy rule $\mu_{ij}\left(x_k\right) = e^{\frac{-\left(x_j(k) - x_{ij}^*\right)^2}{2\sigma_j^2}}$.

This is particularly important for the focal points of rules that might be formed based on the first principle, 1) following expression (13) which might lie too close to each other. The third principle, 3), in fact, leads also to simpler fuzzy rule-base to be formed if compare to other clustering methods such as ART [27], VQ [28] etc. which often require later so called 'pruning' [29].

The rationale of the expression (15) is related to the so called 'one-sigma' condition, $\left|x_j(k) - x_{ij}^*\right| > \sigma_j$ known from the machine learning literature [10]. That is, expression (15) is true when in the rule base there is a fuzzy rule, i such that the input vector of the current data sample, x_k is described in all dimensions, j by at least $e^{-1} \approx 0.36$.

Based on these principles, the following algorithm for on-line learning the antecedents of the fuzzy system based on the data density can be formulated

---

Read data sample *z(k)=[x(k);y(k)]*

$IF\ (k = 1)\ THEN$

//initialization stage//

//initialize the variables for recursive calculation//

$d_j(k)=0; j=[1,n];\ b(k)=0$

// the input part of the first data sample is the focal point of the first cluster (rule)//

$x_1^*(1) \leftarrow x(1)$; $P(z_1^*(k)) \leftarrow 1; R \leftarrow 1$

//form the antecedent part of the first fuzzy rule//

$Rule_1\ IF\ \left(x_{11}\ is\ x_{11}^*\right)\ AND...AND\ \left(x_{1n}\ is\ x_{1n}^*\right)$

$ELSE$

Recursively calculate potential of the current data sample, $P(z(k))$ by (7)-(10);

Update the spread of the clusters (membership functions of the respective fuzzy sets) by (11);

Recursively update the potentials of the existing cluster centers, by (12);

Check

Condition A (A1 or A2);

Condition B where

A1) the point is with ***high*** potential and ***covers new area of data space – expressed*** by (13) or (15);

A2) same as above but expressed by (14);

B) the point ***overlap*** with the previously formed fuzzy rules (16);

$IF\ (A)\ THEN\ \left(x(k)\ is\ new\ focal\ point\right)$

$x_{R+1}^*(k) \leftarrow x(k)\ P(z_{R+1}^*(k)) \leftarrow P(z(k)); R \leftarrow R+1$

$IF\ \left(A\ AND\ B\right)\ THEN\ \left(remove\ nearest\ x_l^*(k)\right)$

$x_l^*(k) \leftarrow x(k)\ P(z_{R+1}^*(k)) \leftarrow P(z(k)); R \leftarrow R+1$

Assign the new point to the nearest cluster.

Repeat until end of data stream

Algorithm Recursive data space partitioning based on the data density

As a result of this data density-based on-line clustering procedure the following fuzzy rule base formed of antecedent parts is generated from the data stream:

$$R_i\ IF\left(x_1 is\ x_{i1}^*\right) AND\left(x_2 is\ x_{i2}^*\right)...\left(x_n is\ x_{in}^*\right); i = [1, R] \tag{17}$$

It can then be either; i) stored and later analyzed by an operator; ii) used for prediction at each time step if combined with consequent part identification as described in [15]; iii) used for classification if combined with consequent part identification as described in [19]; iv) used for clustering the data and various applications e.g. in robotics [7,21].

III. ON-LINE INPUT SELECTION IN SELF-LEARNING TAKAGI-SUGENO FUZZY SYSTEMS

In the previous section, and, to the best of our knowledge, in all previous research in on-line fuzzy systems identification indeed, it was assumed that the dimensionality of the input (features) vector, n is pre-defined for each problem at hand. In this paper we propose an approach that breaks this assumption for Takagi-Sugeno (TS) type fuzzy systems and gradually removes inputs (features) that do not contribute to the output. This is based on the on-line estimation of the sensitivity of the output. Because in TS fuzzy systems the output is locally linear, the sensitivity analysis reduces to analysis of the consequent parameters:

$$R_i\ IF\left(x_1\ is\ x_{i1}^*\right)AND\left(x_2 is\ x_{i2}^*\right)...\left(x_n is\ x_{in}^*\right)$$
$$THEN\left(\ y_i\ = \theta_{i0} + \sum_{j=1}^{n} x_{ij}\theta_{ij}\right); i = [1, R] \tag{18}$$

The overall output of the TS fuzzy system is determined as a weighted average of the outputs of the local linear models [22]:

$$y = \sum_{i=1}^{R} \lambda_i y_i \tag{19}$$

where $\lambda_i = \frac{\prod_{j=1}^{n}\mu_{ij}(x)}{\sum_{l=1}^{R}\prod_{j=1}^{n}\mu_{lj}(x)}$ is the firing strength of the i^{th} fuzzy rule.

The importance of each input (feature) can be evaluated by the ratio of the accumulated sum of the consequent parameters for the specific j^{th} input (feature) in respect to all n inputs (features) [23]:

$$\omega_{ij}(k) = \frac{T_{ij}(k)}{\sum_{r=1}^{n} T_{ir}(k)}; i = [1, R];\ j = [1, n] \tag{20}$$

where $\mathrm{T}_{ij}(k) = \sum_{l=1}^{k}\left|\theta_{ij}(l)\right|$ denotes the accumulated sum of parameter values of the i^{th} rule.

Since the inputs, outputs, and the internal variables of the eTS system can be normalized online as described in [7], they are comparable between each other. The value of the weight can be used for a gradual removal of inputs (features) that contribute little to the overall output (see equation (18)). Then the inputs (features), j^* that do not contribute enough to the output and will be removed from the fuzzy system structure at the next time instant can be determined by:

$$\exists j^*\left|\omega_{ij^*}(k) < \varepsilon_1 \sum_{r=1}^{n} T_{ir}(k) \quad i = [1, R]; j = [1, n]; n \le 10\right.$$
$$\exists j^*\left|\omega_{ij^*}(k) < \varepsilon_1 \max_{r=1}^{n} T_{ir}(k) \quad i = [1, R]; j = [1, n]; n > 10\right. \tag{21}$$

where ε_1 denotes the tolerable minimum weight of an input (feature) – suggested value is 3 to 5%.

The condition is in terms of the proportion which the weight of a certain input (feature) is from the total (sum) or the bigger (maximum) of the accumulated sum of parameters. If this proportion is insignificant (less than ε) then this input (feature) is negligible and can be removed without significantly affecting the output. The two conditions differ, because when the number of inputs (features), n is big the total sum may become too large; at the same time, for a small number of features, the sum gives a better representation than the maximum (an averaging effect). This technique for on-line input (feature) selection is illustrated in section V on real industrial data.

The importance of this technique should not be underestimated, because very often in a real environment there are many measurable variables that influence our output. Selecting most informative inputs (features) is a critical task that is usually associated with the pre-processing stages [10] and is addressed by approaches such as PCA [10], GP [24], etc. These approaches, however, require a batch set of data and a fixed model structure. Relaxing the requirement in terms of the most informative inputs (features) significantly improves the flexibility of the model that is being used and adds to the level of autonomy.

IV. On-Line Fuzzy System Structure Simplification

In section II the automatic design of the fuzzy system antecedents based on the on-line estimation of the data density was described. In section III, a methodology to gradually reduce the number of inputs (features) has been described. Learning parameters of the consequents has been described elsewhere [15,16]. In this section techniques that further contribute to the flexibility of the evolving fuzzy rule-based model will be introduced. They are based on on-line monitoring and analysis of the quality of the rule base through the *age* and *utility* of the fuzzy rules.

Age of the cluster and respectively of the fuzzy rule was first introduced in [25]. Here we introduce a simpler formula:

$$Age_i(k) = k - \frac{\sum_{l=1}^{N_i(k)} I_l}{N_i(k)}; i = [1, R] \tag{22}$$

where I denotes the time index of the moment when the respective data sample was read;

The *Age* indicates how old is the information that supports certain fuzzy rule (data that are associated with certain cluster). One can monitor on-line the *Age* of each fuzzy rule and compare this with the *mean Age* that is determined as $\overline{Age}(k) = \frac{1}{R}\sum_{i=1}^{R} Age_i(k)$ which can be updated on-line. One can use the *Age* of the rule to remove older rules [25] or to detect concept drift which corresponds to the inflexed point of the *Age* curve (the point when the derivative of Age in terms of time index, $\frac{d\left(Age\right)}{dk}$ changes its sign [26]. In Fig 1, the through monitoring the age of the fuzzy rules one can detect two major operating shifts in the process at sample 2250 and sample 2689. Around sample 2250, the 6^{th} rule is activated significantly more frequently than other rules and its aging slower. Around sample 2689, the system switches back to the previous status and the 2nd rule becomes "younger" while the 6th rule gets "older" being activated less.

The *utility* of a fuzzy rule is introduced in [23] to represent the degree of support of a fuzzy rule. While the support of a cluster (respectively fuzzy rule) determined as the number of data samples associated to a certain cluster/rule based on the minimum distance criteria [25] determines a crisp (integer) number of data samples that support certain cluster (respectively fuzzy rule) the *utility* can be seen as a fuzzy measure of the support of a fuzzy rule. It is defined [23] as the accumulated firing strength of the respective fuzzy rule for the span of its life:

$$\eta_i(k) = \frac{\sum_{l=1}^{k} \lambda_l}{k - t_i}; i = [1, R] \tag{23}$$

where t_i denotes the time instant when the i^{th} fuzzy rule has been generated.

Utility, η_i accumulates the weight of the rule contributions to the overall output during the life of the rule (from the current time instant back to the moment when this rule was generated). It is a measure of importance of the respective fuzzy rule comparing to the other rules (comparison is hidden in the relative nature of λ (see equation (18)). In Fig. 2 one can observe that the usage (utility) of some fuzzy rules (e.g. rules 1 and 2) is gradually diminished during the process after sample 500, while the utility of some other rules is increased (e.g. rules 5 and 7).

The evolution of *Age* and *Utility* of specific fuzzy rules throughout the process of on-line design and use of TS fuzzy systems with industrial data is illustrated in Figure 1 and 2. Marginal rules (rules for which $\eta_i < \varepsilon_2$; where ε_2 is the level of *utility* below which it is deemed to be marginal) are being removed (suggested value is 3 to 5%).

Figure 1 Age analysis for Propylene synthesis.

Figure 2 Utility evolution for Propylene synthesis.

V. Experimental Results

A number of experiments were carried out on data streams from several real industrial processes. Each data stream was tested with four different settings of the eTS in order to analyse the effects and the performance of the advanced features introduced in this paper. The first test is based on the basic eTS **without** on-line input selection (OIS) and marginal rules removal (MRR); the second test is of eTS with OIS only; the third test is of eTS with by MRR only; and the last test is of eTS enhanced by both OIS and MRR. The overall performance of the proposed approach was analyzed based on a comparison of these experiments on different industrial data sets. In order for a comparison to demonstrate the general accuracy of the proposed approach, the data sets are also tested with the well established conventional offline fuzzy approaches such as ANFIS [31]. In the test with ANFIS each data set is separated into training set and test set. The training set is used for the offline training and the benchmark tests were carried on the test sets. The three data sets were also tested with a number of established and recent approaches, including a feed-forward - back-propagation neural network, and DENFIS [32]. Note that the proposed system can be implemented as a soft sensor [8] which may also be realized as an embedded system (due to the efficient recursive calculations it is possible to realize the algorithm on chip [30]).

A. Oil refinery case study

This data set contains 447 data samples obtained from a crude oil distillation plant in CEPSA oil refinery located in Santa Cruz, Tenerife, Spain (courtesy of Dr. J. J. Macias Hernandez). The aim is to predict the inflammability point by so called Abel index in the process of production of kerosene. Since eTS can work in real-time in online mode, it is considered a promising tool to monitor the quality [8]. The Abel index can be used to guide the operations of process control in a safe and productive way. Conventionally, this analysis is carried out with the use of laboratory analysis periodically, usually on hourly or daily basis. Predictions were based on hourly average estimation. However, such periodical estimation does not closely correspond to the changes in the dynamic process. As common for the laboratory-based benchmarks, the performance was measured based on the standard deviation over the absolute errors between real outputs and the predictions:

$$e = \sqrt{\left(\frac{1}{N_{val}-1}\right)\sum_{i=1}^{N_{val}-1}(\varepsilon_i - \bar{\varepsilon})^2} \qquad (24)$$

Where e is the measurement; N_{val} is the number of samples during the validation period; ε denotes the absolute error. Normally, e lower than *5* is an acceptable accuracy range [8].

It is demonstrated that the proposed technique can produce acceptable precision in modeling (the numerical results are tabulated in Table I) and can also generate a well interpretable rule-base that can be very helpful to the process operators.

TABLE I
RESULTS FOR PREDICTING QUALITY OF THE CRUDE OIL DISTILLATION

Setting	Rules	inputs	rmse	*e*
Neural Network (**off-line**)	**-**	7	2.87	3.43
ANFIS (**off-line**)	29	7	**2.15**	**2.25**
DENFIS	29	7	2.46	-
eTS without OIS and MRR	5	7	2.29	2.37
eTS with OIS only	9	**5**	2.30	2.38
eTS with MRR	3	7	2.19	2.28
eTS with OIS and MRR	**3**	6	**2.18**	**2.27**

***Final** Rule base in the Abel inflammability test*

*R_1: **IF** (P is 5.4%) **AND** (T^{co} is 323.3 °C) **AND …** **AND** (T^{ne} is 126.8 °C)*

***THEN** (A^1=20.2 + 92.7P + … + 0.12 T^{ne})*

*R_2: **IF** (P is 11.7%) **AND** (T^{co} is 365.0 °C) **AND …** **AND** (T^{ne} is 147.6 °C)*

***THEN** (A^2=42.1 + 63.4P + … + 0.10 T^{ne})*

*R_3: IF (P is 5.4%) **AND** (T^{co} is 335.14 °C) **AND …** **AND** (T^{ne} is 136.1 °C)*

***THEN** (A^3=25.2 + 71.9P + … + 0.19 T^{ne})*

B. Propylene case study

The propylene data set is collected from a chemical distillation process run at The Dow Chemical Co., USA (courtesy of Dr. A. Kordon, [26]). The data set consists of *3000* readings from *23* sensors that are on the plant. They are used to predict the propylene content in the product output from the distillation. Some of the inputs may be irrelevant to the model and thus bring noise. Therefore the input selection is very important task, which is usually done off-line as a part of the pre-processing.

Instead, the procedure proposed in this paper is on-line and leads to an effective selection of the small subset of the inputs. One can see that there are two significant shifts in the operating condition which took place, which tests the online learning ability of the system in a new condition. The results (tabulated in Table II) illustrate the online feature selection; the two most informative inputs were selected after *130* samples have been processed. From Table II it is seen that using all the proposed advanced features of eTS a compact fuzzy model of 7 fuzzy rules and two inputs (fuzzy sets per rule) can be evolved on-line which also provides the best precision. This demonstrates that highly compact, transparent and **interpretable** fuzzy models can be designed from data streams on-line using eTS with the enhancement features as described in this paper.

TABLE II
RESULTS FOR PREDICTING PROPYLENE CONTENT OF DISTILLATION

Method	Rules	inputs	rmse	Correl
NN (**off-line**)	**-**	23	0.11	0.963
ANFIS (**off-line**)	29	23	0.11	0.972
DENFIS	235	23	0.11	0.979
eTS without OIS & MRR	23	23	0.10	0.981
eTS with OIS only	23	**2**	0.09	0.982
eTS with MRR only	14	23	0.12	0.974
eTS with OIS and MRR	**7**	**2**	**0.09**	**0.983**

***Final** Rule-base for Propylene:*

*R_1: **IF** (x_1 is 24.6) **AND** (x_2 is 26.3)*
***THEN** ($\bar{y} = -0.039 + \bar{x}_1 - 0.324\bar{x}_2$)*

*R_2: **IF** (x_1 is 39.0) **AND** (x_2 is43.5)*
***THEN** ($\bar{y} = -0.615 + 4.77\bar{x}_1 - 0.340\bar{x}_2$)*

*R_3: **IF** (x_1 is 46.2) **AND** (x_2 is 49.5)*
***THEN** ($\bar{y} = -0.679 + 1.090\bar{x}_1 + 0.450\bar{x}_2$)*

*R_4: **IF** (x_1 is 45.9) **AND** (x_2 is 49.9)*
***THEN** ($\bar{y} = -1.340 + 5.570\bar{x}_1 - 3.320\bar{x}_2$)*

*R_5: **IF** (x_1 is 36.2) **AND** (x_2 is 43.5)*
***THEN** ($\bar{y} = -0.002 + 0.320\bar{x}_1 - 0.065\bar{x}_2$)*

*R_6 **IF** (x_1 is 31.6) **AND** (x_2 is 38.7)*
***THEN** ($\bar{y} = -0.007 + 0.366\bar{x}_1 - 0.129\bar{x}_2$)*

*R_7 **IF** (x_1 is 40.6) **AND** (x_2 is 39.5)*
***THEN** ($\bar{y} = -0.527 + 0.406\bar{x}_1 - 0.345\bar{x}_2$)*

As shown in Figure 3 the more important inputs quickly become distinct indicated by their importance described by (20).

Figure 3 Importance of inputs - at time instances *k=20 and 120*

C. *NOx emission case study*

The last data set was collected from car engines (courtesy of Dr. E. Lughofer, University of Linz Austria) to estimate the NOx content in the emissions that they produce based on the variables that are easy to measure, such as pressure in the cylinders, engine torque, rotation speed, etc. [20]. In total as much as *180* input variables are considered as potential inputs (these also include the physical variables described in the previous sentence taken at different time instants, i.e. with different time delay). In [20] we used off-line input variable selection method and common knowledge to determine the best *5* inputs. Instead, in this paper we start processing all the data inputs available and select the best subset automatically on-line. When we apply on-line inputs selection only the end result is a fuzzy model with *101* inputs and *36* fuzzy rules; when we combine this with the use of *utility* for removal of marginal fuzzy rules we automatically evolve a fuzzy Takagi-Sugeno model with *7* inputs and *13* fuzzy rules (see all the results in Table III). Moreover, the prediction error is lower with the model that has inputs selected automatically.

The evolution of the *utility* for each fuzzy rule is illustrated in Figure 4. It was generated based on the test keeping all the generated rules for the whole testing period. The *utility* of certain rules become lower over the time.

It is interesting to notice that in Table III, the rule removal helps reducing the finally selected inputs. The reason is that the criteria verifying each input based on the importance of the inputs, ω as described in (21) is checked for each rule one by one, rather than on all rules as a whole. Dropping the marginal rules also leads to a removal of some inputs, and consequently causes the inputs to be dropped out.

TABLE III
RESULTS FOR NOX CAR EMISSION ANALYSIS

Model	Rules	inputs	rmse	Correl
NN (off-line)	-	180	0.15	0.934
DENFIS	113	180	0.17	0.917
ANFIS	32	180	0.15	0.932
eTS without OIS & MRR	22	180	0.17	0.914
eTS with OIS only	36	101	**0.13**	**0.947**
eTS with MRR only	14	180	0.17	0.908
eTS with OIS and MRR	**13**	**7**	0.15	0.935

Figure 4 Evolution of the *utility* of the fuzzy rules at time instances *k=200 and k=500*

VI. ACKNOWLEDGEMENTS

The authors are thankful to Dr. Arthur Kordon from The Dow Chemical, Dr. Jose Macias Hernandez from CEPSA, Tenerife, Spain, and Dr. Edwin Lughofer from Johannes Kepler University of Linz, Austria for providing experimental data.

VII. CONCLUSION AND DISCUSSION

A new approach to fuzzy systems structure learning form data streams is proposed in the paper that takes the recently introduced evolving fuzzy Takagi-Sugeno (eTS) models further. The novelty lies in the combination of techniques that allow for a truly flexible and open (expandable and shrinking) structure of the fuzzy rule base and fuzzy sets to be designed on-line. This includes a new approach for fuzzy rules (respectively clusters) *utility* and *age* – based structure simplification, a new approach for automatic *gradual* selection of most influential inputs (features) based on consequents parameter evaluation and sensitivity analysis, and an enhanced method for antecedents learning by data partitioning, including spreads update. All these new and improved techniques were approbated on a range of industrial-based data sets in Section V and demonstrated that very compact and efficient flexible (with open and adaptive structure) fuzzy systems (in particular of TS type) can be generated from data streams on line (possibly in real time). The evolution of quality estimation parameters of the fuzzy system structure (such as age, utility, spread, number of inputs/features) was illustrated on specific case studies and has been analyzed.

REFERENCES

[1] L. Kuncheva, *Fuzzy Classifiers*, Physica-Verlag, 2000, ISBN 3-7908-1298-6.

[2] D. Nauck, R. Kruse, A Neuro-fuzzy method to learn fuzzy classification rules from data, *Fuzzy Sets and Systems*, Vol. 89, pp. 277-288, 1997.

[3] K. J. Cios, W. Pedricz, R.W. Swinarski (1998) *Data Mining Methods for Knowledge Discovery*, Boston, MA, USA: Kluwer Academic Press.

[4] J. A. Roubos, M. Setnes (2001). Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *IEEE Transactions on fuzzy systems, 9*, 516-524.

[5] R. Babuska, Fuzzy Modeling for Control, Kluwer Academic Publishers, Boston, USA, 1998.

[6] P. Angelov, C. Xydeas, Fuzzy Systems Design: Direct and Indirect Approaches, *Soft Computing*, special issue on New Trends in the Fuzzy Modelling part I: Novel Approaches, vol. 10 (9), July 2006, pp.836-849.

[7] X. Zhou, P. Angelov, "Autonomous Visual Self-localization in Completely Unknown Environment using Evolving Fuzzy Rule-based Classifier", *Proc. IEEE Intern. Conf. on Comput. Intelligence Applic. for Defense and Security,* Honolulu, USA, 1-5 April 2007, pp. 131-138.

[8] J. J. Macias-Hernandez, P. Angelov, X. Zhou, Soft Sensor for Predicting Crude Oil Distillation Side Streams using Takagi-Sugeno Evolving Fuzzy Models, Proc. *2007 IEEE Intern. Conf. on Systems, Man, and Cybernetics*, 7-10 Oct., 2007, Montreal, Canada, ISBN 1-4244-0991-8/07, pp.3305-3310.

[9] A. Elgammal, R. Suraiswami, D. Harwood, and L. Davis, "Background and Foreground modeling using non-parametric Kernel Density Estimation for visual surveillance KDE", *Proc. 2002 IEEE* Vol. 90 (7), pp. 1151 – 1163.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* Heidelberg, Germany: Springer Verlag, 2001.

[11] D. Specht, "A general regression neural network", *IEEE Transactions on Neural Networks*, vol. 2 , No 6, pp. 568-576, 1991.

[12] R. R. Yager, D. P. Filev, "Learning of fuzzy rules by mountain clustering, *Proc. SPIE Conf. on Appl. of Fuzzy Logic Technology*, Boston, MA, USA, pp. 246-254, 1993.

[13] S. L. Chiu, "Fuzzy model identification based on cluster estimation", *J. of Intel. and Fuzzy Syst.* vol. 2, pp. 267-278, 1994.

[14] P. Angelov, "An Approach for Fuzzy Rule-base Adaptation using On-line Clustering", *International Journal of Approximate Reasoning*, Vol. 35 , No 3, March 2004, pp. 275-289.

[15] P. Angelov, D. Filev, "An Approach to On-line Identification of Takagi-Sugeno Fuzzy Models*", IEEE Transactions on System, Man, and Cybernetics, part B - Cybernetics*, vol.34, No1, 2004, pp.484-498. ISSN 1094-6977.

[16] P. Angelov, X. Zhou, "Evolving fuzzy systems from data streams in Real time," In *Proc. 2006 International Symposium on Evolving Fuzzy Systems,* Ambleside, Lake District, UK pp. 29-35, IEEE Press.

[17] P. Angelov, *"Evolving Rule-based Models: A Tool for Design of Flexible Adaptive Systems"*. Heidelberg, Germany: Springer, 2002.

[18] P. Angelov, J. Victor, A. Dourado, D. Filev, On-line evolution of Takagi-Sugeno Fuzzy Models, In *Proc. 2nd IFAC Workshop on Advanced Fuzzy/Neural Control,*16-17 September 2004, Oulu, Finland, pp.67-72.

[19] P. Angelov, X. Zhou, F. Klawonn, "Evolving Fuzzy Rule-based Classifiers", In *Proc. 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing, CIISP 2007,* Honolulu, HI, USA, 1-5 April 2007, pp.220-225.

[20] P. Angelov, E. Lughofer. P. E. Klement, Two Approaches for Data-Driven Design of Evolving Fuzzy Systems: eTS and FLEXFIS, *The 2005 North American Fuzzy Information Processing Society Ann. Conf.*, June 21-25 2005, Ann Arbor, Michigan, USA, pp.31-35.

[21] P. Angelov, R. Ramezani, X. Zhou, Autonomous Novelty Detection and Object Tracking in Video Streams using Evolving Clustering and Takagi-Sugeno type Neuro-Fuzzy System, 2008 World Congress on Comput. Intel., WCCI-2008, Hong Kong, June 2008, to appear.

[22] R. Yager, D. Filev, R. R. Yager, D. P. Filev, *Essentials of Fuzzy Modeling and Control*, NY: John Wiley, 1994.

[23] P. Angelov, Machine Learning (Collaborative Systems) UK patent application, GB-0621734.3, 1 November 2006.

[24] G. Smits, A. Kordon, E. Jordaan . C. Vladislavleva, and M. Kotanchek, Variable Selection in Industrial Data Sets Using Pareto Genetic Programming, In: *Yu T., R. Riolo, B. Worzel (Eds): Genetic Programming Theory & Practice III*. Springer, NY, pp. 79-92, 2006.

[25] P. Angelov, D. Filev, Simpl_eTS: A Simplified Method for Learning Evolving Takagi-Sugeno Fuzzy Models, *The 2005 IEEE International Conference on Fuzzy Systems FUZZ-IEEE*, Reno, Las Vegas, USA, 22-25 May 2005, pp.1068-1073.

[26] P. Angelov, A. Kordon, X. Zhou, Evolving Fuzzy Inferential Sensors for Process Industry, *3rd Int. Workshop on Genetic & Evolving Fuzzy Syst.,* 4-7 March, 2008, Witten-Bomerholz, Germany, to appear.

[27] G. A. Carpenter and S. Grossberg, "Adaptive Resonance Theory (ART)," in The Handbook of Brain Theory and Neural Networks, M. A. Arbib, Ed. Cambridge, MA: MIT Press, 1995, pp. 79–82.

[28] R. Gray, "Vector quantization," IEEE ASSP Magazine, pp. 4–29, 1984.

[29] E. Lughofer, E. Huellermeier, and E. Klement, "Improving the interpretability of data-driven evolving fuzzy systems," in Proceedings of EUSFLAT 2005, Barcelona, Spain, 2005, pp. 28–33.

[30] M. Everett, P. Angelov, *EvoMap: On-Chip Implementation of Intelligent Information Modelling using EVOlving MAPping*, Tech. Report, 2005, Lancaster University, Lancaster, UK, pp.1-15.

[31] J. S. R. Jang, ANFIS: Adaptive Network-based Fuzzy Inference Systems, IEEE Trans. on Systems, Man & Cybernetics, part B – Cybernetics, vol. 23, No 3, 1993, pp. 665 - 685.

[32] N. Kasabov and Q. Song, DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction, IEEE Trans. on Fuzzy Systems, vol.10, No. 2, 2002, pp. 144 - 154.