A Simple Fuzzy Rule-based System through Vector Membership and Kernel-based Granulation

Plamen Angelov
Department of Communication Systems, Infolab21,
Lancaster University, Lancaster, LA1 4WA, UK
e-mail: p.angelov@lancaster.ac.uk

Ronald Yager
Machine Intelligence Institute, Iona College
New Rochelle, New York, USA
e-mail: yager@panix.com

***Abstract*—It is widely recognized that the human reasoning can be approximated by fuzzy rule-based (FRB) systems which can be seen as one of the basic frameworks for representation of intelligent systems. During the last quarter of a century two particular types of FRB systems, namely Zadeh-Mamdani (ZM) and Takagi-Sugeno (TS) dominated the field. In this paper we propose an alternative type which is simpler and more intuitive while preserving the advantages of its predecessors, such as flexibility, modularity, human-intelligibility. The newly proposed concept of vector membership (VM) and kernel-based granulation (KG) of complex systems (respectively their mathematical descriptions) we see as the next, more efficient form of system modelling that is widely applicable to a plethora of applications ranging from time-series prediction, clustering, classification, control, decision support systems to other problems where conventional fuzzy rule-based systems are used. The proposed simple FRB based on VM and KG are non-parametric and fully represent the real data. Contrast this to the mere approximation of the real data distributions that is provided by Gaussian (scalar), triangular, trapezoidal etc. parametric types of membership functions that are used in currently existing types of FRB (ZM and TS). Note that even probabilistic models that are usually based on Gaussian distributions or a mixture of Gaussians or other parametric representations provide only an approximation of the real data distribution (it should be noted that particle filters are perhaps the only form of non-parametric representation that is similar in this sense to the newly proposed simple FRB with VM and KG, but they are computationally cumbersome with exponentially growing complexity). The main contribution of the proposed simple FRB with VM and KG is that while preserving all the advantages of 'traditional' FRB systems they avoid the well known problems related to (multiple scalar) membership functions definition, identification and update. They fully take into account and exactly represent the spatial distribution and similarity of all the real data by proposing an innovative and much simplified form of the antecedent part. At the same time, transformations to the 'traditional' (ZM and TS) fuzzy sets expressed by parametric membership functions per variable are also possible. In papers that will follow we will demonstrate on practical examples (including classification, prediction, decision support and other classes of problems) the benefits of this scheme.**

Keywords- fuzzy rule-based systems, Zadeh-Mamdani and Takagi-Sugeno fuzzy systems, memebership functions, granulation, kernel-based representation

I. INTRODUCTION

During the last quarter of a century fuzzy rule-based (FRB) systems emerged and are now widely accepted as the dominant and main framework to represent the intelligent systems (systems that have elements of human reasoning and certain level of intelligence). Two particular types of FRB systems, namely Zadeh-Mamdani (ZM) [1,2] and Takagi-Sugeno (TS) [3] dominated the field. In this paper we propose an alternative type which is simpler and more intuitive while preserving the advantages of its predecessors, such as flexibility, modularity, human-intelligibility.

The proposed new simpler type of FRB systems is based on vector membership (whereas the traditional concept of fuzzy sets and systems [1-4] is based on a scalar membership type function (per variable) which usually takes the form of a Gaussian, triangular, trapezoidal, bell type function etc. [1-4]. The proposed concept is based on a kernel type representation of the input-output mapping and in this respect it has strong links to support vector machines, Parzen windows [5], neural networks (NN) as well as case-base reasoning [6]. Similarly to multi-model systems concept and, in particular, TS type FRB [1], radial-basis function type NN the proposed concept also starts with decomposing complex non-linear, non-stationary problems into a set of loosely connected local simpler (linear, singleton, exponential, etc.) models which are later aggregated (similarly to TS FRB, in a fuzzy way.

The proposed simple FRB system with VM and KG can be seen as an extension of the well known concept of the case-base reasoning [6] but with a much more sophisticated mathematical underpinning. From the point of view of the fuzzy set theory, the proposed concept provides an innovative representation of the very 'roots' on which fuzzy sets concept is based [1]. The intention of the authors is to simplify the FRB systems and this to remove the problems related to the scalar membership functions definition and parametric representation.

The degree of membership to a parameterized scalar (per individual variable) fuzzy set is replaced with the degree of similarity between different data samples/ measurements in a vector form which makes it much easier to compare different objects in a qualitative form, rank ordered etc. Moreover, there is no need to specify the prototype/focal point/centre of a

978-1-4244-5164-7/10/$26.00 ©2010 IEEE

fuzzy set. This will be determined based on the data distribution from the data in an objective manner.

For example, it is easier to evaluate the degree of similarity between two objects (e.g. patients) rather than to specify the similarity per feature and moreover to properly parameterize and aggregate the scalar degrees of similarity.

The similarity/dissimilarity is closely linked with the notion of distance. In the proposed concept there is no specific request to use (the most widely used) Euclidean distance; it works equally fine with Mahalonobis, cosine or any other suitable for the particular application distance measure. It should be noted that the choice of the distance measure may have an effect on the aggregation result in the vector representation, but this is also the case with the traditional FRB (ZM and TS) and other forms of system representation such as SVM, NN, etc.

The proposed concept touches the very foundations of complex systems representation and thus its application domain ranges from simple clustering-based techniques for pattern recognition, image segmentation, vector quantization etc. to more general modeling, prognostics, classification and time-series prediction problems in various application areas, e.g. intelligent sensors, mobile robotics, advanced manufacturing processes, sensor networks, etc. The applications are deliberately left for the future publications to keep the core idea clear.

II. THE CONCEPT OF THE PROPOSED METHOD

Let us start with comparing the two widely used traditional FRB systems. We notice some striking similarities - both types share exactly the same antecedent part and only differ by their consequents part:

ZM: $Rule^i : IF\left(\left(x_1 is\ LT_1^1\right)\ AND\ \left(x_2 is\ LT_2^1\right)...AND\left(x_n is\ LT_n^1\right)\right)$
$THEN\left(y^i\ is\ LT_{n+1}^i\right);\ i=[1,N]$

TS: $Rule^i : IF\left(\left(x_1 is\ LT^1\right)\ AND\ \left(x_2 is\ LT^2\right)...AND\left(x_n is\ LT^n\right)\right)$ (1)
$THEN\left(y^i = A^i x_e\right);\ i=[1,N]$

where $x_j; j=[1,n]$ denotes the j^{th} input variable; $LT_j^i; i=[1,N]; j=[1,n]$ denotes the j^{th} linguistic term (e.g. *Small, Medium, Large* etc.) for the i^{th} fuzzy rule; y denotes the output variable; A denotes the vector of parameters, $A^i=\left[a_0^i \quad a_1^i \quad \ldots \quad a_n^i\right]^T$; $x_e^T=[1,x^T]$ denotes the extended inputs vector.

The consequent part of the TS type FRB has a crisp, functional form while the consequent part of ZM type FRB has a fuzzy sets – based linguistic form. The aggregation of the contributions of all fuzzy rules to the overall output is usually done in both types by a fuzzily weighted centre-of-gravity type operator (sometimes, e.g. in classification problems they use 'winner takes all' or its variations such as 'few winners take all' aggregations).

The proposed simplified FRB is based on a very powerful and intrinsically generic multi-input-multi-output (MIMO) modelling framework that covers various types of systems, including but not limited to FRB and NN, see Figure 1. We assume a complex, generally non-linear, non-stationary, non-deterministic system that can only be described and observed by its input and output vectors, $x=[x_1,x_2,...,x_n]^T$ and $y^i=[y_1^i,y_2^i,...,y_m^i]$ respectively, Figure 1. Note that this NN form of representation is significantly simpler than the NN form of representation of TS FRB.

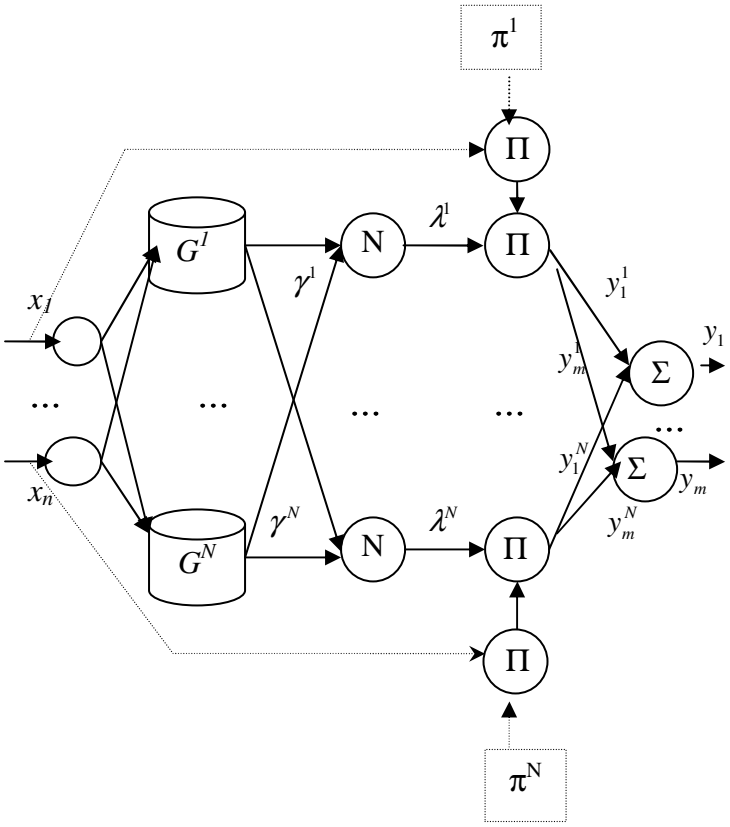

Figure 1. A simple vector membership and grannulation based fuzzy rule-based system in a form of a neural network

The proposed simple FRB describes the input-output dependence based on a summarized history of observation of input-output pairs, $z_j=[x_j^T;y_j^T]^T$, *j=1,2,...k-1* and current, *k* inputs, x_k^T only. The dimensions of the input-output data vector z_j is *(n+m)*: *n* dimensions of the inputs and m dimensions of the outputs. The proposed concept is of a kernel-based granulation of the overall data into granules, *G* which are then associated to respective fuzzy rules ***directly*** in a vector form (over vectors z_j and x). The degree of similarity between a ***current*** data vector and ***all*** previous data samples is calculated thus decomposing the input-output data space, $z \in R^k$ into Granules, *G*. This can be done using a computationally efficient recursive procedure [7]**.** The granules are very similar to the clusters that can be used for design of TS fuzzy sets [5,

10]. The main difference is that the clusters are parameterized through their centre (usually, the mean) and parameters such as the spread of the Gaussian, left and right end of the respective triangular, trapezoidal etc. ***scalar*** membership functions and are, in fact, an ***approximate*** representation of the ***real*** data distribution while the granules are **parameter-free** and use **all the real data**.

A granule is described by the statement

$$\left(z \; is \; like \; \Im^i\right) \tag{2}$$

where the $\Im^i \Big| \quad \Im \in R^{n+m}; \quad i=[1,N]$ denotes a granule (subset of real data with similar properties); $\Im^i \subset Z$; Z denotes the total set of real data samples; $Z=\bigcup_{i=1}^{N}\Im^i$; $\bigcap_{i=1}^{N}\Im^i=\varnothing$.

The degree of membership to a granule is measured by the normalized (using, for example, the centre of gravity aggregation [4]) similarity between a particular data sample, z_j and ***all*** the granules, G:

$$\lambda^i=\frac{\gamma^i}{\sum_{j=1}^{N}\gamma^j}; i=[1,N] \tag{3}$$

where γ^j is the *granule's local* density of the i^{th} Granule for a particular data sample, z_k (if a time series, this can be the data sample taken at the k^{th} time instant, but generally the index k denotes the k-th data sample and not necessarily time).

Note that the vector degree of membership to a granule is in respect to all points/data samples that are associated to that Granule and is not related to a particular point (focal point, centre, apex) nor it requires a radius, contour etc. In other words – it is non-parametric.

The local density of the Granule can be represented by a suitable kernel, e.g. Gaussian, Cauchy, etc.:

$$\gamma_k^i=K\left(d_{ik}\right) \tag{4}$$

where $K(.)$ denotes the kernel function; d_{ik} is the distance between the current sample, z_k and ***all*** the other samples of that Granule

The distance can be represented by any suitable form, e.g. Euclidean: $\left[d_{ik}^2\right]_{Euclidean}=\left\|z_k-z^i\right\|^2$, cosine $\left[d_{ik}^2\right]_{\cos ine}=1-\frac{\sum_{j=1}^{Ni} z_{kj}z_{ij}}{\sqrt{\sum_{j=1}^{Ni} z_{kj}^2 \sum_{j=1}^{Ni} z_{ij}^2}}$,

etc. N_i denotes the number of samples in the i^{th} Granule.

If use Cauchy type of kernel which is particularly suitable for recursive calculations [7,8] we arrive at:

$$K\left(d_{ik}\right)=\frac{1}{1+\sum_{j=1}^{Ni}\frac{d_{ijk}^2}{k-1}}=\frac{k-1}{k+\sum_{j=1}^{Ni}d_{ijk}^2-1} \tag{5}$$

where z^i denotes ***all*** data samples that belong to the Granule, $\Im^i$ that is $z^i: \; \forall z \left| z \in \Im^i \right.$;

The n-dimensional projection of a Granule which concerns only the inputs can then be defined as:

$$\left(x \; is \; like \; \aleph^i\right) \tag{1a}$$

where the degree of membership to a Granule is measured by the normalized similarity between a particular input data vector, x_j and all the input vectors from all the Granules, $\aleph^i \Big| \quad \aleph \in R^n; \quad i=[1,N]$

If a scalar membership function is preferred in some applications, we suggest using the following approach – breaking down the range of the scalar variable of interest, x_i into a number (Int) of small ***equal*** intervals – the number of these intervals will determine the level of discretisation of the scalar membership function. Count (possibly recursively) the number of data samples for which the variable x_i has value lying between the border values of the respective interval, e.g. $x(k) \in \left[\overline{x}_i^j ; \underline{x}_i^j\right]$; where $\overline{x}_i^j$ denotes the upper boundary of the interval and $\underline{x}_i^j$ denotes the lower boundary respectively. Let us denote the number of data samples/measurements that fall into the j^{th} interval by $n_{\Delta j}$ which should also be a function of time, k. Then the scalar membership function can be represented by the ratio:

$$\mu_i^j(k)=\frac{n_{\Delta j}(k)}{\max_{j=1}^{Int} n_{\Delta j}(k)} \; ; j=1,2,\ldots, Int \tag{6}$$

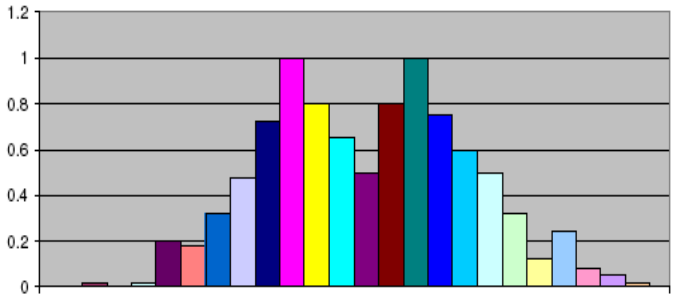

Figure 2. An illustration of the scalar membership function of non-parametric type which fully and exactly reflects the data distribution (on the vertical axis the degree of membership is denoted; on the horizontal axis the (scalar) variable intervals are positioned.

This formula will provide a 2D visualization of the scalar data distribution of the variable x_i. The value of *1* will be assigned to intervals with most data samples and respectively smaller values to other intervals. The size of the interval can be changed (which will provide a different resolution of representation of the membership function, but once chosen it should be the ***same*** for all intervals considered for this variable for a particular moment of time, k.

Note that this 2-D visualization is not necessary for the method but is simply provided as a possible illustration and a link to the traditional fuzzy sets which have scalar 2-D membership functions. It also demonstrates that the proposed approach is much closer to the real data distribution than (usually regular – Gaussian, triangular, trapezoidal etc.) traditional scalar parameterized membership functions and parametersied distributions.

The proposed simple FRB system looks very similar to the ZM and TS FRB in that it also has its antecedent and consequents part and can be described by a set of simple linguistic fuzzy rules:

$$Rule^i: IF\left(x\ is\ like\ \aleph^1\right)\ THEN\left(sub-mo\ del^i\right) \tag{7}$$

where $Rule^i$ denotes the i^{th} fuzzy rule; $i=[1,N]$; N is the overall number of fuzzy rules; $\left(x\ is\ like\ \aleph^i\right)$ denotes the n-dimensional projection of the i^{th} granule (sub-set of the real input data); *sub-model*^{i} is the *Local Model* of the i^{th} fuzzy rule, which is usually (but not necessarily) linear, e.g.

$$y^i = x_e^T \pi^i \tag{8a}$$

where $x_e^T = [1, x^T]$; $\pi^i = \begin{bmatrix} a_{01}^i & a_{02}^i & \cdots & a_{0m}^i \\ a_{11}^i & a_{12}^i & \cdots & a_{1m}^i \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}^i & a_{n2}^i & \cdots & a_{nm}^i \end{bmatrix}$ are the parameters of the m *Local Models*; or singleton-based *Local Models*:

$$y^i = A^i \tag{8b}$$

where $A^i = \left[a_{01}^i \quad a_{02}^i \quad \cdots \quad a_{0m}^i\right]^T$ are the *Local Model* parameters;

The overall output, y can be formed as a collection of fuzzily combined multiple *Local Models* (simpler sub-systems), y_i:

$$y = \sum_{i=1}^{N} \lambda^i y^i \tag{9}$$

where y^i represents the output of the i^{th} local model.

The main innovation is in the much simplified antecedents part which is base don granules and a vector membership represented through the (Cauchy) kernels.

III. Design of Simplified FRB Systems

In this section the design/identification of the newly proposed simplified FRB with VM and KG will be described. It should be stressed that the newly proposed simplified FRB systems are not necessarily linked or limited by the off-line or on-line or evolving type of system identification [7,8]. It can be realized in any of these types including based on expert knowledge, see Figure 3.

The design of any system has two main aspects; i) system structure identification, and ii) parameter identification/tuning. The problem of system structure identification is often left to the choice of the system designer. This problem was paid much more attention since Mountain clustering [9] was proposed to be used to automatically solve the problem of FRB systems design. Later, the concept of *system structure evolution* [10] further developed this design technique. The problem of parameter tuning/optimization has been traditionally more widely developed [11].

If we consider a classifier as an example (we have stressed that the proposed new concept has much wider implications to predictors, controllers etc.) it can be developed in the following simple way. During the training phase (which can be on-line) each data sample has a set of features (input vector) and labels (output vector). A simple FRB classifier can be designed assuming a number of Granules equal to the number of classes (it should be stressed that this is not necessarily required and the number of Granules in general can be larger than the number of classes, but not smaller). If assume the same number of Granules as the number of classes the design procedure reduces to assigning a data sample to the respective Granule of that class and updating the density, γ^i by a recursive version of (5) (the full derivation and proof that (10) is equivalent to (5) is provided in [7,8]):

$$K\left(d_{ik}\right) = \frac{k-1}{(k-1)\left(z_k^T z_k + 1\right) - 2\alpha_k + \beta_k} \tag{10}$$

where $\alpha_k = z_k^T \xi_k;\ \xi_k = \xi_{k-1} + z_k;\ \ \beta_k = \beta_{k-1} + \left\|z_{k-1}\right\|^2$

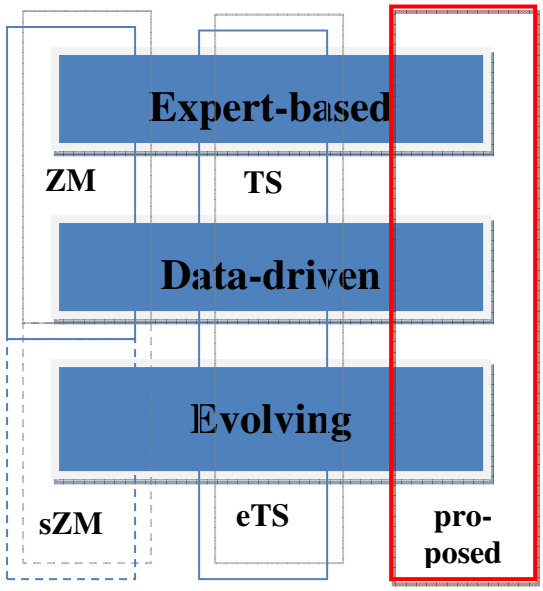

Figure 3. The newly propsoed simplified FRB system in the context of ZM and TS and modes of operation; sZM denotes simplified ZM model (when singletons are used in the consequents instead of fuzzy sets).

In this way a classifier of the following vector form is designed autonomously from the data:

$$Rule^{i}: \quad \begin{array}{l} IF \ \left(x \ is \ similar \ to \ G^{i}\right) \\ THEN \ \left(Class^{i}\right) \end{array} \quad i=[1,N] \qquad (11)$$

A. System Structure Identification

Ideally, granulation should satisfy the following main principles [8, 11]

a) good generalization and summarization of data by Granules with high density (low variance of the data inside the granule), and
b) good coverage of the entire data space by expanding area of interpolation.

To allow a continuous evolution of the structure, Granules (rules/neurons) are being removed if they stop to be used actively or get older (if a *concept shift* is detected in the data stream). The *concept drift* [12] is linked to a *smooth sliding* of the data distribution through the data/feature space from one granule to another. The *drift* is closely related to the time-space representation of the data streams. While the concept of (data) density is represented in the data space domain, *drift* and *shift* are concepts in the joint data-time space domain.

The proposed method includes self-monitoring and self-regulation of the quality of the generated Granules. In an on-line and evolving version of this approach [7,8] only Granule prototypes and the global mean value are necessary to be memorized (these are *N+1* values of dimension *(n+m)*), while ***all*** the other data points are discarded. The question arises 'how well these centers represent the data that were discarded from the memory?'. One way to address this issue is by monitoring properties of the Granules that are formed. These include their ddf, local *density*, mean, standard deviation, *support*, *age*, *utility*. These are described in more details elsewhere [8,10].

B. Self-Learning Method

The second phase involves learning the consequent part's parameters which represents parameter identification. Once the antecedent part of the fuzzy rule-based model is determined and fixed the identification of parameters of the consequent part, π^i can be solved as a recursive least square (RLS) estimation problem [11]. The real-time algorithm must perform both tasks (data partitioning and parameter estimation) at the same time instant (per data point) for a time significantly shorter than the sampling period. In this way, the antecedent part of the rules can be determined in a fully unsupervised way, while the consequent part requires a supervised feedback. The supervision is in the form of error feedback which guarantees optimality (subject to fixed rule base/neural network structure) of the parameters of the consequent part.

The overall output of the newly proposed simplified FRB system can be given in a vector form as follows:

$$y = \psi^T \theta \qquad (12)$$

where $\theta = \left[\left(\pi^1\right)^T, \left(\pi^2\right)^T, ..., \left(\pi^N\right)^T\right]^T$ is a vector formed by the sub-system parameters; $\psi = [\lambda^1 x_e^T, \lambda^2 x_e^T, ..., \lambda^N x_e^T]^T$ is a vector of the inputs that are weighted by the normalized activation levels of the rules, λ^i , *i=[1,N]* for the linear consequents, (8a) and $\psi = [\lambda^1, \lambda^2, ..., \lambda^N]^T$ for the singleton type consequents, (8b).

For a given data point, x_k the optimal in LS sense solution $\hat{\theta}_k$ that minimizes the following cost function:

$$\left(Y - \Psi^T \theta\right)^T \left(Y - \Psi^T \theta\right) \rightarrow \min \qquad (13)$$

can be found applying weighted RLS, wRLS (which is described in more detail in [7,8]):

$$\hat{\theta}_k = \hat{\theta}_{k-1} + C_k \psi_k (y_k - \psi_k^T \hat{\theta}_{k-1}) \qquad (14)$$

$$C_k = C_{k-1} - \frac{C_{k-1} \psi_k \psi_k^T C_{k-1}}{1 + \psi_k^T C_{k-1} \psi_k} \qquad (15)$$

where $\hat{\theta}_1 = 0$; C is a $Nn \times Nn$ co-variance matrix; $C_1 = \Omega I$; Ω is a large positive number; I is the identity matrix; $k=2,3,\ldots$

wRLS is fuzzily weighted through the activation levels and is not the conventional weighted RLS which is directly applicable under the assumption that the model (12) has a fixed structure. Under this assumption the optimization problem (13) is linear in parameters. The concept of *evolving* systems assumes a *gradually evolving* model structure. As a result, the activation level of the fuzzy rules, λ^i will change. These changes (even infrequent and gradual in the sense that only one out of *N* rules is affected) have retrospective effect in the sense that they affect previously calculated activation levels, λ_j^i (*i=[1,N]*; *j=[1,k-1]*). Local wRLS is significantly less affected by this disturbance to the theoretical optimality of the RLS condition. In addition it is significantly less computationally complex.

IV. CONCLUSIONS

In this paper we proposed an alternative type of FRB which goes further in the conceptual and computational simplification while preserving the best features (flexibility, richness combined with simplicity, modularity) of its predecessors (ZM and TS type FRB). The newly proposed concept is seen as the next, more efficient form of system modelling applicable to time-series prediction, clustering, classification, control, decision support systems and other problems where conventional fuzzy rule-based systems are used. The simplified FRB through VM and KG has a non-parametric form that ***fully*** reflects the real data (instead of attempting to approximate them with parametric functions, e.g. Gaussian, triangular, trapezoidal etc. as the conventional systems do). The main contribution of the proposed simplified FRB systems are that they avoid the well known problems related to membership functions definition, identification and update while preserving all the advantages of the traditional

FRB systems. They ***fully*** and ***exactly*** take into account the spatial distribution and similarity of ***all*** the ***real*** data by proposing an innovative and much simplified form of the antecedent part. At the same time, transformations to the 'traditional' fuzzy sets expressed by parametric membership functions per variable are still possible through projections per scalar variables. In papers that will follow we will demonstrate on practical examples (including classification, prediction, decision support and other classes of problems) the benefits of this scheme.

ACKNOWLEDGMENT (IN MEMORIAM)

This paper is devoted to the late Professor Abe Mamdani who gave u s one of the most popular forms of fuzzy rule-based systems. If we succeeded to achieve a simpler and more efficient form of these systems we will be happy to have taken a little bit further his great pioneering result.

REFERENCES

[1] L. A. Zadeh, Outline of a New Approach to Analysis of Complex Systems and Decision Processes, *IEEE Transactions on Systems, Man and Cybernetics*, vol.1, pp.28-44, 1973.

[2] E. H. Mamdani, S. Assilian, An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, *International Journal of Man-Machine Studies*, vol.7, pp.1-13, 1975.

[3] T. Takagi, M. Sugeno, Fuzzy identification of systems and its application to modeling and control, *IEEE Trans. on Syst., Man & Cybernetics*, vol. 15, pp. 116-132, 1985.

[4] R. Yager and D. Filev, Essentials of Fuzzy Modeling and Control (John Wiley & Sons, 1994).

[5] R. O. Duda, P. E. Hart, and D.G. Stork. *Pattern Classification - Second Edition*. Wiley- Interscience, Chichester, West Sussex, England, 2000.

[6] I. Watson, Case-based reasoning is a methodology not a technology, *Knowledge-Based Systems*, vol. 12 (5-6), pp. 303-308, Oct. 1999.

[7] P. Angelov, D. Filev, An approach to on-line identification of evolving Takagi-Sugeno models, *IEEE Trans. on Syst., Man and Cybernetics, part B*, vol.34 (1), pp. 484-498, 2004.

[8] P. Angelov, Machine Learning (Collaborative Systems) World patent, priority date, 1 Nov. 2006; international filing date 23 Oct. 2007.

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

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

[11] T. Kailath, A. H. Sayed, B. Hassibi, *Linear Estimation*, Upper Saddle Rive, NJ, USA: Prentice Hall, 2000.

[12] G. Widmer and M. Kubat. Learning in the presence of concept drift and hidden contexts. Machine Learning, vol. 23 (1):69–101, 1996.