2011 IEEE International Conference on Fuzzy Systems
June 27-30, 2011, Taipei, Taiwan

Fuzzy Granular Evolving Modeling for Time Series Prediction

Daniel Leite and Fernando Gomide
School of Electrical and Computer Engineering
University of Campinas
Campinas SP, Brazil
Email: {danfl7,gomide}@dca.fee.unicamp.br

Rosangela Ballini
Institute of Economics
University of Campinas
Campinas SP, Brazil
Email: ballini@eco.unicamp.br

Pyramo Costa
Graduate Program in Electrical Engineering
Pontifical Catholic University of Minas Gerais
Belo Horizonte MG, Brazil
Email: pyramo@pucminas.br

***Abstract*—Modeling large volumes of flowing data from complex systems motivates rethinking several aspects of the machine learning theory. Data stream mining is concerned with extracting structured knowledge from spatio-temporally correlated data. A profusion of systems and algorithms devoted to this end has been constructed under the conceptual framework of granular computing. This paper outlines a fuzzy set based granular evolving modeling - FBeM - approach for learning from imprecise data. Granulation arises because modeling uncertain data dispenses attention to details. The evolving aspect is fundamental to account endless flows of nonstationary data and structural adaptation of models. Experiments with classic Box-Jenkins and Mackey-Glass benchmarks as well as with actual Global40 bond data suggest that the FBeM approach outperforms alternative approaches.**

***Index Terms*—Data Stream, Evolving System, Granular Computing, Online Learning, Time Series.**

I. Introduction

Continuous data stream processing has become a task of primary importance mainly due to the emergence of industrial sensor networks and small scale computing devices that produce massive amounts of data from their environments. In online settings, large unbounded datasets stream in high frequency and bring uncertainty in their instances. Data streams demand fast single-scan-through-the-data algorithms to identifying and storing essential information. Rethinking many existing data mining and machine learning techniques is necessary to consider structural adaptation of information systems based on sequences of never-seen-before instances.

Recent research on evolving granular systems [1]-[7] lays emphasis on granulated views of detailed data and computing with granules coarser than the data in order to simplify complex real-world problems and provide low cost solutions. As stated by Zadeh [8] and Yao [9], granular computing exploits the tolerance for imprecision, uncertainty and partial truth to achieve tractability, robustness and better rapport with reality. The flexibility of handling dynamics within granular framework enables us to describe granules in different application domains without deep knowledge about the underlying problem. Narrow time and space constraints of online environment as well as understandability requirements inspire granulated views of the data and computing at coarser granularities.

Fuzzy set based evolving modeling FBeM approach belongs to the framework of granular computing because it uses fuzzy-type information granules to construct granular maps that associate input granular datum to output granular datum. Fuzzy granules warrant the generality of the structure of the data and give algorithms with simple math and rules describing their meaning. Building fuzzy sets from uncertain data is the FBeM strategy for lodging noise and disturbances. Basically, FBeM systems look to streaming data under different resolutions and decide when to shift back or forth between simpler or more detailed granularities. Structured representation of data streams by means of a set of fuzzy rules that carry the very essence of the information mirrored in the data is a wealthy contribution.

FBeM benefits from fuzzy granular objects to summarize information-in-motion and assist decision making. Models in FBeM are created on the fly and evolved as requested by the data stream. Supported data sources include sensors, web traffic, audio and video, financial data, climate data, etc. The FBeM learning algorithm creates and fosters granules recursively, steered by the information flow. Eventually, the quotient granular structure may be optimized, refined or coarsed, agreeing with inter-granule relationships.

Structurally, FBeM models combine functional and linguistic fuzzy systems to provide singular and granular approximations of nonstationary functions. Functional fuzzy systems are generally more precise whereas linguistic fuzzy systems are more interpretable. Accuracy and interpretability require tradeoffs and one usually prevails over the other. By combining functional and linguistic systems into a single modeling framework, $\cup$-closure for short, FBeM takes advantage of both systems simultaneously. At the practical level, experts usually prefer that online systems give approximated results as well as tolerance bounds on the approximations.

This paper addresses time series prediction. Examples on predicting the Box-Jenkins, Mackey-Glass and Global40 time series emphasize the complementarity of the functional and linguistic parts of FBeM and illustrate the usefulness of the approach. The FBeM predictor makes no specific assumption about the properties of the data sources in experiments, but rather lets the data stream guide learning freely.

The remainder of this paper is organized as follows. Section II addresses the granular fuzzy modeling approach, the FBeM structure and its characteristics. The associated data-stream-driven recursive learning algorithm is detailed in Section III. Section IV presents results of time series predictions with

FBeM playing the role of an evolving predictor. Section V concludes the paper and suggests issues for further investigation.

II. Fuzzy Set Based Evolving Modeling

FBeM is an evolving modeling approach that produces higher level information granules based on more detailed streaming data and recursive learning algorithm. A $\cup$-closure granular structure ensues from more specific local models. Incremental learning procedures cast the FBeM structure to track new concepts, cope with uncertainty, and provide singular and granular approximations of nonlinear functions. FBeM addresses the problem of unbounded databases and the scalability issue. It deals with computationally hard problems outer approximating the solution.

An FBeM model conveys a set of If-Then fuzzy rules extracted from the data. The set of rules means a granular representation of a complex system. Learning in FBeM starts from scratch, i.e., no granules and rules need to be preconceived nor the amount of evolved granules ceiled. Granules and rules are created and adapted on the fly, dynamically, steered by the behavior of the process function over time. Whenever data instances arrive, a decision mechanism is trigged and either granules and rules are inserted into the FBeM structure or parameters of existing granules are adapted.

FBeM rules manage information granules. For each granule there exists a corresponding rule. They are gradually evolved over time. Experts may wish to provide a verbal description about the process using their intuition and experience. Evolving fuzzy modeling, Fig. 1, supports both, learning from data streams and learning from experience.

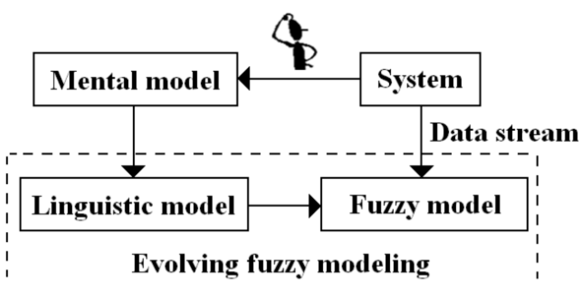

Fig. 1. Evolving fuzzy modeling

In FBeM models, rules R^i governing information granules γ^i are of the type

$$\begin{array}{l} \text{IF } (x_1 \text{ is } A_1^i) \text{ AND...AND } (x_j \text{ is } A_j^i) \text{ AND...AND } (x_n \text{ is } A_n^i) \\ \text{THEN} \quad (y_1 \text{ is } B_1^i) \quad \text{AND} \quad \bar{y}_1 = p_1^i(x_j \forall j) \quad \text{AND} \\ \qquad ... \\ \qquad (y_k \text{ is } B_k^i) \quad \text{AND} \quad \bar{y}_k = p_k^i(x_j \forall j) \quad \text{AND} \\ \qquad ... \\ \qquad \underbrace{(y_m \text{ is } B_m^i)}_{\text{linguistic}} \quad \text{AND} \quad \underbrace{\bar{y}_m = p_m^i(x_j \forall j)}_{\text{functional}} \quad , \end{array}$$

where x_j and y_k are variables of the data stream $(x,y)^{[h]}$, $h = 1, ...$; A_j^i and B_k^i are membership functions built in light of the data being available; p_k^i are approximation polynomials. The collection of rules R^i, $i = 1, ..., c$, forms the rule base. Rules born and grow on-demand whenever the structure of the data calls for improvements in the current model. It is worth noting that an FBeM rule combines both, linguistic and functional consequents. The linguistic part of the consequent outer approximates functions and provides interpretability of the results. The functional part of the consequent offers singular approximation of functions and precision. Using this structure, FBeM takes advantage of both, linguistic and functional systems, within a single modeling framework.

A. Rule Antecedent

Granulation of data into fuzzy objects A_j, $j = 1, ..., n$, can be based on grid, tree or scatter partitioning. Grid partition divides attributes in finer parts $\{A_j^1, ..., A_j^c\} \prec A_j$ of equivalent size (top-down approach). Low-level granules represent more specific concepts; $\prec$ is a fine relation. A fixed fuzzy grid is easily interpretable but suffers with concept change, learning in unknown domain, and coexistence of different granularities in the data stream. Tree partition results from cuts of larger granules. Trees are more flexible than grids since sub-trees A_j^i, refinements of A_j, can be sliced independently. Trees may experience the problem of excessive partitioning due to large amount of features and concept drift. Scatter fuzzy partitioning uses subsets A_j^i which can be extended to fuzzy hyperboxes in a product space by means of alpha level sets. The scattering process clusters the data into information granules when appropriate. Granules are positioned at arbitrary locations into the product space. An aspect to be taken into account with scattering-type granulation refers to searching for a suitable amount of partitions, their positions and sizes. Figure 2 illustrates the different types of data granulation.

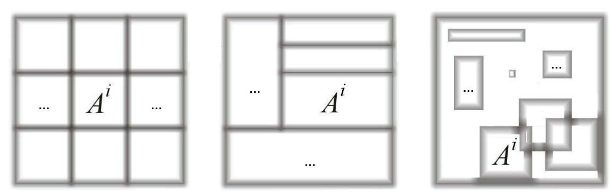

Fig. 2. Grid, tree and scattering granulation

In this work, the FBeM system is equipped with scattering-type granulation mechanism. Accommodating data into conveniently placed and sized granules leaves substantial flexibility for incremental recursive adaptation and freedom in choosing the internal representation of the granules. Online environment claims for opportune creation and rearrangement of fuzzy structural objects from time to time. Essentially, FBeM seizes Gaussian membership functions as formal granular objects to wrap uncertainty in the data of streams. Gaussians are quite easily converted to hyperboxes by means of alpha level sets.

A Gaussian fuzzy subset $A_j^i = \mathcal{G}(\mu_j^i, \sigma_j^i)$ is characterized by a modal value μ_j^i and spread σ_j^i. Features that make the Gaussian representation suitable include: (*i*) easiness of acquiring the necessary parameters. The modal and spread values are captured straightforwardly from a data stream; (*ii*) infinite support does not ignore data. Because data ranges are unknown before learning, the support of Gaussians extends across the whole domain; (*iii*) smoothness and continuously differentiable surface. On representing uncertainty, Kreinovich [10] under certain assumptions has shown that Gaussian functions are the most adequate choice for representing imprecise measurements.

B. Rule Consequent

Consequent of FBeM rules joins functional and linguistic fuzzy information to approximate actual system outputs and provide tolerance bounds on the approximation. The functional part of the consequent p_k^i comprises singular local functions whereas the linguistic part depicts corresponding information granules B_k^i occurring along the domain of output variables k.

Here we assume affine local functions of the type

$$p_k^i = a_{0k}^i + \sum_{j=1}^{n} a_{jk}^i x_j \quad (1)$$

for all instances measured from the process function f and that activate the fuzzy hyperbox delineated by the granule γ^i. In general, each p_k^i can be of different type and is not required to be linear. For instance, higher order polynomials could be used to approximate f. However, the number of coefficients to be identified increases substantially, especially when the number or variables n is large. FBeM approximands p_k^i estimate f within the domain of granule γ^i. The recursive least mean square (RLMS) algorithm is used to determine the corresponding coefficients a_{jk}^i.

Gaussian representation allows all granules to overlap. Therefore, each rule in FBeM contributes to the system output. FBeM singular output is determined as a weighted mean value over all rules,

$$p_k = \frac{\sum_{i=1}^{c} min(A_1^i, ..., A_n^i) p_k^i}{\sum_{i=1}^{c} min(A_1^i, ..., A_n^i)}. \quad (2)$$

This assure smooth transition between overlapped functions.

Similarly to the approach for embodying antecedent of rules A_j^i, consequents of rules B_k^i benefit from scattering-based granulation and fuzzy hyperboxes to cluster output data streams skillfully. We assume Gaussian fuzzy subsets $B_k^i = \mathcal{G}(\mu_k^i, \sigma_k^i)$ to assemble granular objects in the output space by the same motivations previously described for A_j^i.

Granular output such as B_k^i may enrich decision making and sometimes provides more useful information than the specific numerical output p_k. Whilst being very specific using p_k we risky being incorrect, being unspecific from B_k^i we turn ourselves assure with a certain confidence of being correct. Sacrificing accuracy pays the price of the guarantee of correctness. Information granules tend to reflect the essence of the structure of the underlying data stream and emphasize the interpretability of the result.

III. Recursive Online Learning

FBeM learns online from a stream of instances $(x, y)^{[h]}$, $h = 1, ...$, where $y^{[h]}$ is known given $x^{[h]}$ or will be known at some latter step. Each pair (x, y) is an observation of the target function f. When f changes with the time we say that the function is nonstationary. Modeling nonstationary functions requires tracking time-varying functions $f^{[h]}$. Data streams require choosing a modeling framework and designing recursive algorithm to decide when and how to proceed structural and parametric adaptation of models.

The learning procedure to evolve fuzzy granular systems FBeM can be summarized as follows:

```
Begin
  Do
    1: Input a new instance (x, y)^[h], h = 1, ...
    2: Accommodate possible new information
      2.1: Create a new information granule and a rule
      2.2: Adapt existing granules and rules
    3: Discard instance (x, y)^[h]
    4: Optimize the quotient granular structure
End
```

Steps 1 and 3 of the learning procedure emphasize the essence of data stream driven algorithms, i.e., instances are read and discarded one at a time. Historical data are dispensable and evolution stands continuously. Granular systems evolve whenever new information appears in the data, step 2. When a new instance does not fit current knowledge, the procedure is to create a new information granule and a rule governing the granule, step 2.1. Conversely, if a new instance fits current knowledge, the procedure adapts existing granules and rules, step 2.2. Eventually, the quotient structure may be optimized, coarsed or refined, according with inter-granules relationships, step 4. Next sections detail the procedure.

A. Creating Rules

In FBeM no rule does necessarily exist before learning starts. Rules are born and evolve when data are input. A new granule γ^{c+1} is created adding a rule R^{c+1} to the current collection of rules $R = \{R^1, ..., R^i, ..., R^c\}$. Whenever the current collection of rules is not sufficiently activated by an instance $x^{[h]}$ so to support a confident output, a new granule is created. FBeM assumes that the instance $x^{[h]}$ brings new information about the process.

Formally, let $\rho \in [0, 1]$ be a threshold value that determines whether to create or adapt rules. If

$$min(A_1^i, ..., A_n^i) \leq \rho \quad \forall i, \quad (3)$$

then the FBeM structure is extended. Note that if ρ is set to 0, then the FBeM system is structurally stable and unable to capture eventual concept shift. Conversely, if ρ equals 1, the FBeM system creates a rule for each new instance, which is not practical. Life-long adaptability is reached through maintaining a condition between the extremes, see Fig. 3.

The role of ρ is crucial in determining the granularity of FBeM models. Choices of ρ impact model accuracy and model transparency, e.g, shifting back and forth between values of ρ

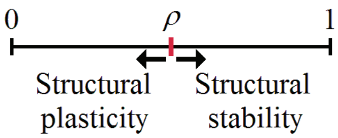

Fig. 3. Plasticity-stability tradeoff and the role ρ in FBeM systems

results in different granulated views of the same process in different levels of detail.

A new granule γ^{c+1} is initially represented by membership functions, A_j^{c+1} and B_k^{c+1}, with parameters

$$\mu_j^{c+1} = x_j^{[h]}, \quad \mu_k^{c+1} = y_k^{[h]} \text{ and } \quad \sigma_j^{c+1} = \sigma_k^{c+1} = 1/2\pi, \tag{4}$$

i.e., the Stigler approach for standard Gaussian functions [11]. The coefficients of local-valued polynomials p_k^{c+1} are set to

$$a_{0k}^{c+1} = y_k^{[h]} \quad a_{jk}^{c+1} = 0, \; j \neq 0. \tag{5}$$

With that initial parameterization, preference is given to the design of granules balanced along all dimensions rather than granules with unbalanced geometry. Therefore, FBeM follows the principle of the balanced information granularity [12] and tends to give more specific rules in the sense of Yager [13].

B. Adapting Rules

Rule adaptation consists in (*i*) expand or contract objects A_j^i and B_k^i to accommodate new data; (*ii*) move granules γ^i toward denser regions of data over the input and output domains; and simultaneously (*iii*) adjust coefficients of local approximation functions p_k^i.

A rule R^i is adapted always it is sufficiently activated by a data instance $x^{[h]}$ according to

$$min(A_1^i, ..., A_n^i) > \rho. \tag{6}$$

Geometrically, the instance rests into a region highly influenced by the granule γ^i. To include $x^{[h]}$, FBeM updates the modal value and the spread of membership functions A_j^i recursively as follows:

$$\mu_j^i(\text{new}) = \frac{(\varpi - 1)\mu_j^i(\text{old}) + x_j}{\varpi^i} \tag{7}$$

$$\sigma_j^i(\text{new}) = \frac{(\varpi^i - 1)}{\varpi^i}\sigma_j^i(\text{old}) + \frac{1}{(\varpi^i - 1)}(x_j - \mu_j^i(\text{new}))^2 \tag{8}$$

where ϖ^i stands for the number of times that the granule γ^i has been activated by the data stream. Notice that values are recursively computed and they do not need data accumulation. Also, only the most active rule for $x^{[h]}$ is chosen for adaptation.

Adaptation of fuzzy sets of rules consequents B_k^i uses output data $y_k^{[h]}$. Polynomial coefficients a_{jk}^i are updated based on the standard RLMS algorithm and takes advantage of the new instance that has activated γ^i. Storage of a number of recent instances may be useful to guide alternative coefficient identification algorithms, e.g., data chunks oriented algorithms. However, it comes with some additional cost concerning memory and processing time.

C. Setting the Granularity

The granularity threshold ρ takes values in the unit interval according to prediction errors. Levels of activation of rules for a given input $x^{[h]}$ are compared to the value of $\rho^{[h]}$ to define either structural or parametric changes of FBeM systems. Values of ρ influence the granularity and understandability of local models. In the most general case, FBeM starts learning with an empty rule base and ignorance about the task and the properties of the data. Therefore, it is fair to initialize ρ in an intermediate condition to permit structural stability and plasticity equally. We use $\rho^{[0]} = 0.5$ as default initial value.

Let E be the maximum square error between predictions $p_k(x^{[h]})$ and actual values $y_k^{[h]}$, then

$$e_k = (y_k^{[h]} - p_k(x^{[h]}))^2, \; k = 1, ..., m, \tag{9}$$

and

$$E = max(e_1, ..., e_k, ..., e_m). \tag{10}$$

Assume E_D is the desired prediction error, and let ρ learn values for itself from

$$\rho(\text{new}) = \rho(\text{old}) + \alpha(E_D - E), \tag{11}$$

where α is a learning rate. Decision makers have authority over the value of E_D and may wish it to be zero. Clearly, very small values of E_D conduct ρ to 0 and lead data overfitting. Practice suggests trade-offs to attain acceptable approximation error and useful granular abstractions and compactness of the original dataset into interpretable rules. Recursive adaptation of the granularity of FBeM systems alleviates guesses on how fast and how often the structure of the data changes.

D. Coarsening the Quotient Structure

As the data arrive and granules are identified, relationships between pairs of granules may be strong enough to justify forming a larger and more abstract granule that inherits the essence and nature of the lower level smaller granules. Quantitative analysis of inter-granule relations requires a suitable metric to measure the distance between uncertain objects.

With Gaussian membership functions as internal representatives of granules, the expected distance between uncertain instances belonging to two different granules, say γ^{i_1} and γ^{i_2}, may be given by:

$$D(\gamma^{i_1}, \gamma^{i_2}) = \frac{1}{n}\sum_{j=1}^{n} ||\mu_j^{i_1} - \mu_j^{i_2}||^2 + \sigma_j^{i_1} + \sigma_j^{i_2} - 2\sqrt{\sigma_j^{i_1}\sigma_j^{i_2}}. \tag{12}$$

This distance has demonstrated to be of fast calculation and more accurate than e.g. distance between means in L^p spaces [14]. The relation considers uncertain data and the specificity of the information, which is in turn inversely proportional to the spread. FBeM combines granules using the lowest entry of $D(.)$ for any pair of granules of the current collection and a decision criterion, which may be either based on a threshold value Δ, or expert judgment in respect of the convenience of the combination.

A new granule γ^i, coarsening of γ^{i_1} and γ^{i_2}, is constituted by Gaussian membership functions with modal value

$$\mu_j^i = \frac{\frac{\sigma_j^{i_1}}{\sigma_j^{i_2}}\mu_j^{i_1} + \frac{\sigma_j^{i_2}}{\sigma_j^{i_1}}\mu_j^{i_2}}{\frac{\sigma_j^{i_1}}{\sigma_j^{i_2}} + \frac{\sigma_j^{i_2}}{\sigma_j^{i_1}}},\ j = 1, ..., n, \tag{13}$$

and spread

$$\sigma_j^i = \sigma_j^{i_1} + \sigma_j^{i_2},\ j = 1, ..., n. \tag{14}$$

These are heuristic relations that basically take into account the proportion of uncertainty of each granule being combined to determinate the location and size of the new granule. The same coarsening relations hold for output variables k. The coefficients of the new local polynomial are

$$a_{jk}^i = \frac{1}{2}(a_{jk}^{i_1} + a_{jk}^{i_2}),\ j = 0, ..., n. \tag{15}$$

Naturally, combining granules reduces the number of rules in FBeM systems and redundancy.

E. Removing Granules

A granule should be removed from the FBeM structure if it seems to be inconsistent with the current concept. Common removing strategies either (*i*) retire stale granules by age, (*ii*) exclude the weakest granules based on error values or (*iii*) delete the most inactive granules. In FBeM, we opt by the strategy of deleting inactive granules. Stale granules may still be useful in the current environment while weak granules are attempted to be invigorated through adapting parameters of singular and granular functions.

FBeM granules are deleted when they become inactive during a number of processing steps, h_r. If the application requires memorization of rare events, or if cyclical drifts are expected, then it may be the case to let all granules live forever. Removing inactive granules periodically helps to keep the rule base updated.

IV. Time Series Prediction

Observing past outcomes of a system to estimate its future behavior: time series prediction is based on the idea that the series carry the potential information needed to predict their future values. Analyzing data produced by actual phenomena can give good insights into the phenomena itself and knowledge about the rules underlying the data.

Chaotic time series are commonly found in natural phenomena [15]-[16]. Models of chaotic time series are useful both to test some hypothesis or theory about the generating phenomena to predict future values of the series, and to decide on mechanisms to control future values. In this avenue, researchers have relied on methods based mainly on looking at trends, moving averages, and graphical patterns to perform predictions. Most of these are linear approaches, e.g. the Box-Jenkins method, linear regression, Kalman filtering, and suffer of some shortcomings [17]. More recently, predicting approaches into the realm of soft computing, e.g. wavelet networks, fuzzy and neuro-fuzzy predictors, have emerged. These approaches have advantages over the traditional statistical ones [18] because generally they are nonlinear in nature thus approximating complex dynamics more easily.

Although classical neural networks and fuzzy systems can approximate any nonlinear continuous function, they usually demand high quality training data and time-consuming off-line learning. Learning system models from data streams in online mode is a challenging task for most of the classical statistical and emergent computing techniques. Therefore, there exists substantial room for further developments of predictors able to perform online learning using data streams.

A. Problem Statement

Forward prediction of a time series can be stated as follows. Given a finite sequence of observations of a discrete time series, namely, $x^{[h]}$, $x^{[h-1]}$, ..., $x^{[h-M]}$, find the continuation $x^{[h+1]}$, $x^{[h+2]}$, ... This involves setting a scalar M and encountering a function f such that a future value, e.g. $x^{[h+1]}$, can be predicted by

$$\hat{x}^{[h+1]} = f(x^{[h]}, x^{[h-1]}, ..., x^{[h-M]}). \tag{16}$$

This is equivalent to modeling the time series as

$$x^{[h+1]} = f(x^{[h]}, x^{[h-1]}, ..., x^{[h-M]}) + \zeta^{[h+1]} \tag{17}$$

with $\zeta^{[h+1]}$ being a white noise process. If the statistics of the time series are non-Gaussian or the time series is the result of some nonlinear operation, the function f is nonlinear. When the characteristics of the time series change with time we say f is nonstationary. Nonlinear nonstationary functions require that online adaptive models identify time varying relations $f^{[h]}$ to perform prediction.

B. Box-Jenkins Gas Furnace

The Box and Jenkins gas furnace benchmark data [19] consist of 9-second-sampling measurements recorded from a combustion process of an oxygen-methane gas mixture. The proportion of gases was randomly changed during the process. The original dataset consists of 296 time indexed instances of the gas flow rate into the furnace $x^{[h]}$ and the concentration of carbon dioxide $y^{[h]}$ emitted.

The task of the FBeM system is to identify the level of carbon dioxide at instant h, $y^{[h]}$, using the preview observation

$y^{[h-1]}$ and the gas rate four steps before $x^{[h-4]}$. This premise prevails in literature works due to positive correlation between these state variables. FBeM scans the data only once to build the system structure and adjust parameters. This is to emulate a data stream. The rule base is initially empty.

Starting parameters to be set include $\alpha = .05$, $E_D = .04$, $\Delta = .08$ and $h_r = 300$. These are values that have particularly worked well for a range of problems. In this experiment we prioritized the compactness and performance of the model.

Training and testing are performed concomitantly from $h = 5$ to 296. First, the estimation $y^{[h]}$ for a given instance $(x^{[h-4]}, y^{[h-1]})$ is provided by the system (testing). Then, parametric and structural adaptation is carried out when necessary (training). Training is necessary whenever an instance carries new information significantly mismatching the current knowledge. Sample-per-sample testing-before-training approach portrays the true online data stream context.

Performance evaluation is made based on the root mean square error:

$$RMSE = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(y^{[h]} - p^{[h]})^2} \qquad (18)$$

and the non-dimensional error index:

$$NDEI = \frac{RMSE}{std(y^{[h]}\forall h)}, \qquad (19)$$

which ponders the $RMSE$ by the inverse of the standard deviation of the underlying data.

Table I summarizes the results. It emphasizes the FBeM performance against alternative models in identifying levels of carbon dioxide smoked out of the furnace. The results of Table I show that FBeM is the most accurate model according to the $RMSE$ and $NDEI$ indices. FBeM reaches the best performance using an average of 2.61 ± 0.55 rules during learning with a maximum of 3 rules, as shown in Fig. 4. The algorithm does not take advantage from large amount of local processing units as evidenced in the figure, but from a combination of ingredients concerning with structural assumptions, peculiarities of the learning algorithm, and fuzzy granular framework to achieve the performance.

TABLE I
GAS FURNACE - PREDICTION PERFORMANCE

Model	Ref.	Rules	$RMSE$	$NDEI$
ARMA	[19]	–	0.8426	3.8687
TS	[20]	2	0.2608	1.1973
ANFIS	[21]	25	0.0854	0.3923
HyFIS*	[22]	15	0.0648	0.2975
eTS*	[23]	5	0.0490	0.3057
Simp-eTS*	[23]	3	0.0484	0.3004
FBeM*	This paper	3	0.0421	0.1932

* Online model

Figure 4 illustrates FBeM one-step singular and granular predictions for the gas furnace problem. The granular prediction $\mu \pm \sigma$ associated to the more meaningful singular prediction p is an important information which may help decision making giving an idea about a range of values around the prediction.

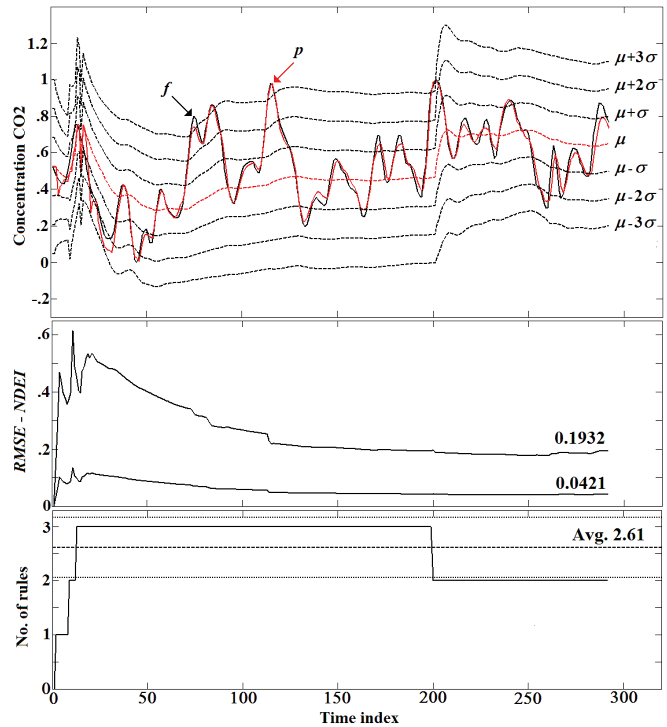

Fig. 4. FBeM singular and granular one-step prediction, and evolution of the error indices and number of rules for the Box and Jenkins gas furnace data

Rules of particular interest can be displayed at any time. For example, a rule at $h = 296$ is:

R^i: IF $x^{[h-4]}$ is $\mathcal{G}(.4192, .1269)$ AND $y^{[h-1]}$ is $\mathcal{G}(.6523, .1516)$ THEN $y^{[h]}$ is $\mathcal{G}(.6503, .1497)$ AND $p^i = .2514 - .0656x^{[h-4]} + .8128y^{[h-1]}$.

C. Mackey-Glass

The Mackey-Glass equation described by:

$$\frac{dx}{dt} = \frac{Ax^{[t-\tau]}}{1 + (x^{[t-\tau]})^C} - Bx^{[t]}, \;\; A, B, C > 0, \qquad (20)$$

is a time delay differential equation that behaves chaotically or periodically depending on the values of its parameters and on the time delay τ. The equation may represent a model of feedback control systems.

The task of FBeM is constructing a function:

$$x^{[t+\xi]} = p(x^{[t]}, x^{[t-\Delta]}, ..., x^{[t-D\Delta]}). \qquad (21)$$

Similar to many studies on this series we admit $\xi = 85$, $\Delta = 6$, $D = 3$; and $A = 0.2$, $B = 0.1$, $C = 10$, $\tau = 17$ for the parameters of the Mackey-Glass equation to generate a state vector. Data are presented sequentially to the FBeM system one at a time. The system starts learning from scratch, with no rules nor pre-training.

To evaluate the effect of different parameterizations, we conduct two experiments. Firstly, FBeM^1 prioritizes a compact structure and adopts $\alpha = .05$, $E_D = .06$, $\Delta = .2$ and $h_r = 2000$. Second, FBeM^2 focuses high accuracy at the price of a larger structure and employs $\alpha = .1$, $E_D = .01$, $\Delta = .1$ and $h_r = 8000$. We resort to the sample-per-sample testing-before-training approach from $h = 105$ to $h = 11898$. Table II shows the performance of the FBeM models and that of other online models for the Mackey-Glass prediction problem.

Table II shows that FBeM^1 is supported by a relatively compact structure, 3.23 ± 0.56 rules with a maximum of 5 rules, similar to eTS, IBeM and xTS, to give competitive results in terms of the $RMSE$ and $NDEI$ indices. Figure 5 details the results of this experiment. The figure illustrates the singular prediction p of the Mackey-Glass series provided by FBeM^1, and the evolution of the number of rules and error indices during the learning steps. Granular prediction is omitted for clarity. Lastly, FBeM^2 makes use of 27.87 ± 4.11 rules with a maximum of 33 rules to sponsor the smallest prediction error and overcome the performance of the remaining approaches.

In case equation f represents the production of white blood cells to defend the human body against pathogen, as in Glass and Mackey [29], FBeM granular prediction may represent, for example, the range of normal activity of hematopoietic stem cells. The effectiveness of the approach in predicting chaotic time series without prior knowledge about the data is clearly verified in this experiment.

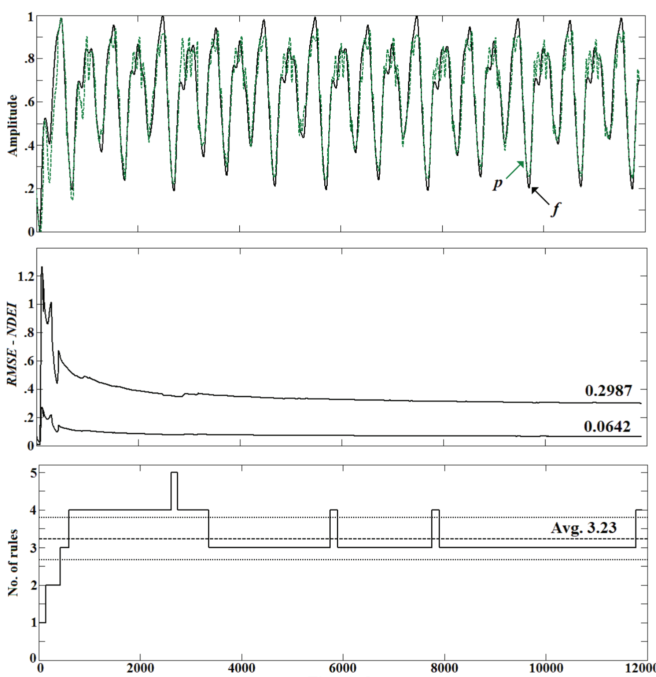

Fig. 5. FBeM singular one-step prediction and evolution of the number of rules and error indices for the Mackey-Glass time series

TABLE II
MACKEY-GLASS - PREDICTION PERFORMANCE

Model	Ref.	Rules	$RMSE$	$NDEI$
EFuNN*	[24]	193	0.0822	0.4010
RAN	[25]	113	0.0802	0.3730
eTS*	[26]	9	0.0799	0.3720
IBeM*	[7]	5	0.0769	0.3577
xTS*	[26]	10	0.0711	0.3310
FBeM^1*	This paper	5	0.0642	0.2987
DENFIS*	[27]	58	0.0593	0.2760
Neural Gas	[28]	1000	0.0133	0.0620
IBeM*	[7]	98	0.0126	0.0586
FBeM^2*	This paper	33	0.0122	0.0568

* Online model

D. Brazilian Global40

Currently, the most negotiated Brazilian sovereign bonds are the Global 2040, which were issued in 2000, with 40 years maturity. Price movements define the bond rentability: higher prices imply lower rentability. Thus, if investors believe that the bond price will increase and the rentability decrease, they should buy the bond before this expected movement. On the other hand, if the investors believe the bond price will decrease and the rentability increase, they should behave contrariwise. Predicting bond price movements is a crux issue from the point of view of the investors on their portfolio decisions.

Variable selection may enhance the prediction ability of FBeM as well as provide useful insights about the data source. Time series carry a limited number of indispensable variables, while the remainders tend to confuse the underlying predictor. The input variables of FBeM may be M consecutive measurements of the time series or we may perform selection of $n < M$ variables, which are not necessarily sequential. In this experiment, variables were chosen based on a partial autocorrelation function of the Global40 bond price, similar to [30]. Coincidently, autocorrelation analysis has suggested $x^{[h-1]}$ and $x^{[h]}$ for one step ahead, $x^{[h+1]}$, prediction.

To evaluate the relative performance of the FBeM predictor, we consider daily bond prices from August 11, 2000 to January 16, 2007, testing-before-training in a sample-per-sample basis, and the following parameters: $\alpha = .03$, $E_D = .01$, $\Delta = .1$ and $h_r = 500$. Table III shows the results on one-step-ahead prediction of the bond price. The result using FBeM makes use of 2.25 ± 1.05 rules, with a high of 4 rules during steps $h = 863$ to $h = 1149$ as portrayed in Fig. 6. The FBeM approach achieved error indices equivalent to the ePL approach employing a structure almost as compact as ePL. However, FBeM offers additional information concerning with

TABLE III
GLOBAL 2040 - PREDICTION PERFORMANCE

Model	Ref.	Rules	RMSE	NDEI
eTS*	[23]	7	0.1362	0.5069
xTS*	[26]	12	0.1355	0.5043
AR	[19]	–	0.0136	0.0506
ePL*	[30]	2	0.0123	0.0458
FBeM*	This paper	4	0.0121	0.0452

* Online model

tolerance bounds around the numerical prediction, as highlights the zoom-in of Fig. 6. Granular output in this case may be read as optimistic and pessimistic bias of the rentability of the Global40 bond.

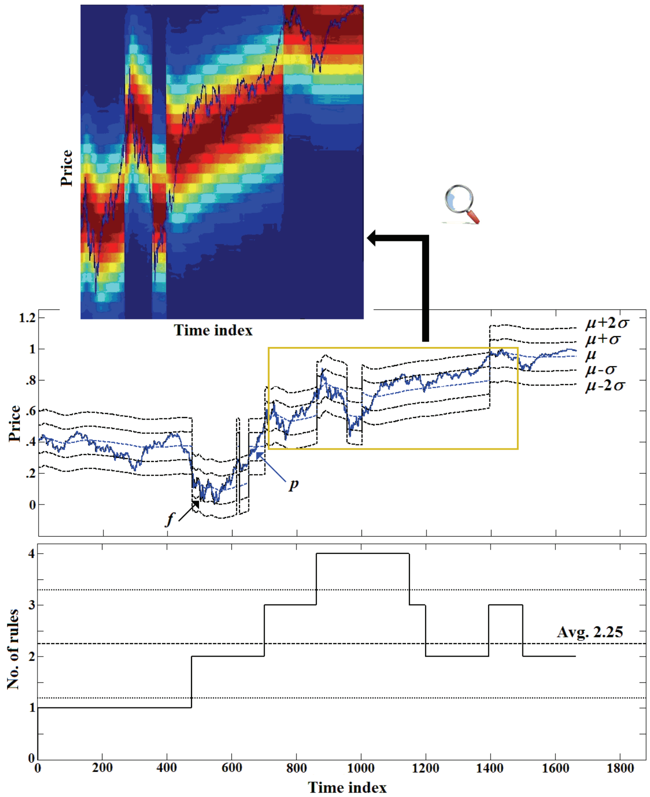

Fig. 6. FBeM singular and granular one-step prediction and evolution of the number of rules for the Global40 bond price

V. Conclusion

This work has suggested fuzzy set based evolving modeling as a framework to learn from online data streams. The FBeM algorithm recursively granulates data instances to output singular and linguistic granular approximations of nonstationary functions. The system combines good accuracy of functional fuzzy models with the advantage of better semantic interpretation of linguistic models. The usefulness of the FBeM approach in prediction was verified using the classic benchmarks Box-Jenkins and Mackey-Glass and actual data of the Global40 time series. Comparisons with alternative approaches have shown the effectiveness of the FBeM modeling approach. Further work shall discuss different forms of manifestation of information granules in data streams and the role FBeM to capture the essence of the information in the data.

Acknowledgment

The first author acknowledges CAPES, Brazilian Ministry of Education, for his fellowship. The second author thanks CNPq, the Brazilian National Research Council, for grant 304596/2009-4. The last author is grateful to the Energy Company of Minas Gerais - CEMIG, Brazil, for grant P&D178.

References

[1] W. Pedrycz, "Evolvable fuzzy systems: some insights and challenges," Evolving Systems, Vol. 1, Issue 2, pp: 73-82, 2010.

[2] A. Bargiela, W. Pedrycz, "Granulation of temporal data: a global view on time series," NAFIPS'03, pp: 191-196, 2003.

[3] P. Angelov, D. Filev, "An approach to online identification of Takagi-Sugeno fuzzy models," IEEE Transactions on Systems, Man, and Cybernetics - Part B, Vol. 34, Issue 1, pp: 484-498, 2004.

[4] D. Leite, F. Gomide, "Evolving linguistic fuzzy models from data streams," To appear in: E. Trillas, P. Bonissone, L. Magdalena, J. Kacprycz (Eds.), Studies in Fuzziness and Soft Computing: A Homage to Abe Mamdani, Springer-Verlag, 15p., 2011.

[5] D. Leite, P. Costa, F. Gomide, "Evolving granular neural network for semi-supervised data stream classification," International Joint Conference on Neural Networks, pp: 1-8, 2010.

[6] D. Leite, P. Costa, F. Gomide, "Interval-based evolving modeling," IEEE Symposium Series on Comp. Intelligence-ESDIS'09, pp: 1-8, 2009.

[7] D. Leite, P. Costa, F. Gomide, "Granular approach for evolving system modeling," In: E. Hullermeier, R. Kruse, F. Hoffmann (Eds.), Lecture Notes in AI, Vol. 6178, pp: 340-349, Springer, 2010.

[8] L. A. Zadeh, "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," Fuzzy Sets and Systems, Vol. 19, pp: 111-127, 1997.

[9] Y. Y. Yao, "Perspectives of granular computing," IEEE International Conference on Granular Computing, pp: 85-90, 2005.

[10] V. Kreinovich, C. Quintana, L. Reznik, "Gaussian membership functions are most adequate in representing uncertainty in measurements," NAFIPS'92, pp: 618-624, 1992.

[11] S. M. Stiegler, "A modest proposal: a new standard for the normal," The American Statistician, Vol. 36, Issue 2, JSTOR, 1982.

[12] A. Bargiela, W. Pedrycz, Granular Computing: An Introduction, Kluwer Academic Publishers - Boston, Dordrecht, London, 452p., 2003.

[13] R. R. Yager, "Measures of specificity over continuous spaces under similarity relations," Fuzzy Sets and Sys., Vol.159, pp: 2193-2210, 2008.

[14] L. Xiao, E. Hung, "An efficient distance calculation method for uncertain objects," IEEE Symposium on Computational Intelligence and Data Mining, pp: 10-17, 2007.

[15] C. Gao, J. Qian, "Evidence of chaotic behavior in noise from industrial process," IEEE Trans. on Signal Processing, Vol.55-6, pp:2877-84, 2007.

[16] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge Univ. Press, 390p., 2004.

[17] S. Haykin, Adaptive Filter Theory, 4^{th} ed., Prentice-Hall, 936p., 2001.

[18] G. S. Maddala, Introduction to Econometrics, 3^{th} edition, Wiley, 664p., 2001.

[19] G. Box, G. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day San Francisco, US, 2^{nd} edition, 1976.

[20] M. Sugeno, T. Yasukawa, "A fuzzy-logic-based approach to qualitative modelling," IEEE Trans. on Fuzzy Systems, Vol.1, Iss.1, pp: 7-31, 1993.

[21] J. Jang, "ANFIS: adaptive-network-based fuzzy inference systems," IEEE Transactions on SMC, Vol. 23, Issue 3, pp: 665-685, 1993.

[22] J. Kim, N. Kasabov, "HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems," Neural Networks, Vol. 12, Issue 9, pp: 1301-1319, 1999.

[23] P. Angelov, D. Filev, "Simpl-eTS: a simplified method for learning evolving Takagi-Sugeno fuzzy models," IEEE International Conference on Fuzzy Systems, pp: 1068-1073, 2005.

[24] N. Kasabov, "Evolving Fuzzy Neural Networks - Algorithms, Applications and Biological Motivation," In: T. Yamakawa, G. Matsumoto (Eds.), Methodologies for the Conception, Design and Application of Soft Computing, World Scientific, pp: 271-274, 1998.

[25] J. Platt, "A resource-allocating network for function interpolation," Neural Computation, Vol. 3, pp: 213-225, 1991.

[26] P. Angelov, X. Zhou, "Evolving fuzzy systems from data streams in real-time," Int. Symp. on Evolving Fuzzy Systems, pp: 29-35, 2006.

[27] N. Kasabov, Q. Song, "DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction," IEEE Transactions on Fuzzy Systems, Vol. 10, Issue 2, pp: 144-154, 2002.

[28] B. Fritzke, "A growing neural gas network learns topologies," Advances in Neural Information Processing Systems, Vol. 7, pp: 625-632, 1995.

[29] M. Mackey, L. Glass, "Oscillation and chaos in physiological control systems," Science, Vol. 197, Issue 4300, pp: 287-289, 1977.

[30] R. Ballini, A. Mendonca, F. Gomide, "Evolving fuzzy modeling of sovereign bonds," The Journal of Financial Decision Making - special issue: The Fuzzy Logic in the Financial Uncertainty, Vol.5, Iss.2, 2009.