The need for qualitative reasoning in fuzzy modeling: robustness and interpretability issues

Raffaella Guglielmann (raffy@imati.cnr.it)
Dipartimento di Matematica - Università di Pavia, via Ferrata 1, 27100 Pavia

Liliana Ironi (ironi@imati.cnr.it)
IMATI - CNR, via Ferrata 1, 27100 Pavia, Italy

Abstract

The most problematic issues in fuzzy modeling of nonlinear system dynamics deal with robustness and interpretability. As for these two issues, traditional data-driven approaches may be affected by serious drawbacks. Especially when the data set is not adequate, they may lead to a model structure that results to be unable to reproduce the system dynamics. Moreover, parameter estimation may lead to a numerically unstable model unless proper regularization strategies are adopted. Also, fuzzy models generated from data only do not guarantee to gain insight into the system: the resulting model structure is often not transparent, and, after their optimization, the model parameters may lead to an incomplete, inconsistent and even indistinguishable fuzzy partition. Herein, we demonstrate that Qualitative Reasoning (QR) plays a crucial role to significantly improve both robustness and interpretability. The method proposed builds both fuzzy partition of input-output variables and the fuzzy rule base from the available physical knowledge only. On the one side, this leads to a clear and neat model structure that does represent the system dynamics, and the parameters of which have a precise physical meaning. Consequently, a substantial improvement of the model generalization and interpretability properties is obtained. On the other side, it allows us to properly constrain the parameter optimization problem, with a consequent gain in numerical stability.

Introduction

Robustness and interpretability are essential prerequisites for a model to be used. The former issue concerns the generalization and stability properties of the identified model, whereas the latter one concerns its transparency and intelligibility as well as the clear and sound physical meaning of the estimated parameters. System modeling goes through two main stages, namely structure identification and parameter optimization, that heavily account for robustness and interpretability. As for input-output approaches, the robustness and interpretability aspects are perhaps the most challenging problems [Girosi et al., 1995, Jin, 2000, Johansen, 1994, Niyogi and Girosi, 1999, Poggio and Smale, 2003, Pomares et al., 2002]. In such approaches, structure identification deals with the reconstruction of functional relationships $f: X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ between the input-output variables from the available data samples only[1]. This problem is usually solved by first selecting an appropriate functional form for $f(\cdot)$, and then by identifying its parameters. The form of $f(\cdot)$ has to be selected in a space that is known to possess good approximation properties. Methods recently proposed for constructing f range from feedforward neural networks [Haykin, 1997] to regularized neural networks [Poggio and Girosi, 1990] to spline models [Wahba, 1990] to fuzzy systems [Wang, 1994]. With the exception of fuzzy systems, these approaches ignore the possible exploitation of available prior knowledge. Their main benefit is essentially that they are able to reconstruct the unknown system dynamics without requiring prior knowledge; but, they may present poor generalization capabilities, ill-posed identification problems, and lack of transparency.

Our interest is focused on Fuzzy Systems (FS) as they have been proved to be excellent candidates for identification purposes [Jang, 1993, Takagi and Sugeno, 1985]: (i) they hold the universal approximation property, (ii) they are able to exploit the qualitative and uncertain a priori knowledge on the system dynamics, which is expressed by inferential linguistic information in the form of IF-THEN rules, and (iii) they are able to handle data samples. In this context, structure identification requires to determine the fuzzy partition of input-output variables, and how many rules must be used to generate the FS. The parameters, which are tuned on the experimental data through optimization procedures, are associated with the membership functions of input-output variables or, in other words, with the locations of their fuzzy partition. In theory, both partitions and inference rules can be derived by the expert knowledge, but such information may be very poor, irregular, and unstructured, and then, in practice, prevents from defining the optimal form of $f(\cdot)$, where by optimal we mean that $f(\cdot)$ is of minimal complexity, but able to capture all of the significant features of the system dynamics. For these reasons, the research efforts turned to the definition of learning methods that automatically generate the fuzzy systems from the data samples only [Wang and Mendel, 1992, Wang, 1994]. Although these methods have been successfully applied to a variety of domains, they are affected by two serious drawbacks: the resulting nonlinear function is not understandable from a physical viewpoint, and it does not guarantee that the generalization property holds unless a large amount of samples is employed. Even when the resulting $f(\cdot)$ is abstracted from the expert knowledge, the same problems may occur since an empirical rather than structural kind of knowledge is mostly given. In such a case, important pieces of information about the system dynamics may be omitted or, in other words, the structure of $f(\cdot)$ may result to be sub-optimal.

Let us observe that for a great deal of dynamical systems from different domains the available structural knowledge is insufficient for the formulation of a quantitative differential model, but does not prevent from formulating a qualitative one. This consideration motivated our work aiming at the definition of a new approach, called FS-QM, to the fuzzy identification of dynamical systems [Bellazzi et al., 1998, Bellazzi

[1] For the sake of simplicity but without loss of generality, we consider here Multiple Input - Single Output systems.

et al., 1999, Bellazzi et al., 2000, Bellazzi et al., 2001]. Its novelty consists in the way the FS is built: both the fuzzy partition and rule base are defined upon the available structural knowledge. FS-QM is applicable whenever the incompleteness of a priori knowledge is such that it allows us (i) to write a QSIM model [Kuipers, 1994], and (ii) to bound the uncertainty on landmark values to a range of numerical values. In outline, the whole range of possible system dynamics, represented and simulated within the QSIM modeling framework, is automatically translated into the fuzzy formalism. The domain of each input/output variable is automatically partitioned into fuzzy sets in accordance with its associated quantity space, and with the prior information on landmark numerical bounds. In other words, the cardinality of the fuzzy partition of a variable and the membership function locations are defined by the cardinality of the set of qualitative values the variable may assume, and by the interval bounds of its landmarks, respectively. Given a landmark-based fuzzy partition and the simulated behavior tree, the generation of the Fuzzy Rule Base (FRB) is straightforward derived by mapping each behavior of the input/output variables into a set of rules, where each rule describes a transition from a qualitative state to the next one. The mathematical interpretation of the FRB explicitly initializes $f(\cdot)$ which is then refined through parameter identification from the data samples. Let us remark that the idea of exploiting QR techniques for system identification is not new within the QR community but most of the work done addresses the problem of differential modeling [Capelo et al., 1998, Kay et al., 2000, Bradley et al., 2001].

Herein, the emphasis is on recent results dealing with robustness and interpretability aspects of the model generated within the FS-QM framework. Robustness is affected by factors related to the FRB, such as its completeness and consistency, and by the possible numerical instability of the parameter optimization procedures. These factors may also make interpretability problematic. As for interpretability, other issues deal with both the completeness and distinguishability and shape of the membership functions that characterize the variable fuzzy partitions. We will demonstrate that the qualitative model underlying the FS allows us to tackle and solve the problems above in a neat and efficient way.

The results herein reported deals with the identification of Thiamine kinetics in the intestine tissue. The classical differential approach turned out to be inapplicable for the incompleteness of the available knowledge and for the difficulty of gathering an adequate number of experimental data. This latter cause was also responsible for the failure of conventional fuzzy approaches [Bellazzi et al., 2001].

Background

The reconstruction of nonlinear system dynamics from data may be seen as a problem of modeling nonlinear discrete-time dynamical systems. Among the possible schemes to describe the system dynamics of the output variable y [Ljung, 1987], let us consider the following one:

$$y_k = f(\underline{x}_{k-1}, \underline{\theta}) + \epsilon_k \tag{1}$$

where the output y measured at time k is a function of a n-dimensional regressor vector $\underline{x}$, which includes both output and input variables, measured at time $k-1$. The function $f(\cdot)$ is unknown and expresses the functional relationship between the output and the input vector, $\underline{\theta}$ is the vector of parameters, and the terms ϵ_k's, independent, zero mean random variables, account for the measurement errors. Then, the fuzzy modeling problem consists in finding a continuous function approximator $\tilde{f}$ of f within a proper class of FS's. As for function approximation two different classes have been widely used, the Takagi-Sugeno [Takagi and Sugeno, 1985], and the Center Average Defuzzifier [Mamdani, 1974]. Herein, we consider the latter class. The knowledge on the relations between input-output variables is expressed through IF-THEN rules:

IF x_1 is F_1 and ... and x_n is F_n THEN y is F_y.

The antecedents x_i are the input vector components, the consequent y is the output, F_i and F_y are fuzzy sets characterized by a membership function $\mu_F : \mathbb{R} \rightarrow [0, 1]$. In accordance with the interpretation of fuzzy operators in [Wang, 1994], M rules are mathematically interpreted by the following fuzzy system:

$$\tilde{f}(\underline{x}, \underline{\theta}) = \frac{\sum_{j=1}^{M} \hat{y}_j [\prod_{i=1}^{n} \mu_i^j(x_i, \underline{\hat{\theta}}_i^j)]}{\sum_{j=1}^{M} [\prod_{i=1}^{n} \mu_i^j(x_i, \underline{\hat{\theta}}_i^j)]} \tag{2}$$

where the parameter $\hat{y}_j$ is the center of the μ that characterizes F_y in the j-th rule; μ_i^j, that depends on the parameter vector $\underline{\hat{\theta}}_i^j$, characterizes the fuzzy set associated with the variable x_i in the j-th rule; the vector $\underline{\theta}$ includes all of the parameters. But, whatever class is selected, the problem of system modeling goes through two sub-problems that should be separately solved to make the modeled system behavior easily interpretable and transparent:

1. *Structural identification*:
   (a) For each variable, define its fuzzy partition, i.e. the μ's that define the fuzzy values it may assume. The locations of the μ's initialize $\underline{\theta}$;
   (b) define the optimal number of rules M, and the rules;
   (c) mathematically interpret the rules.
2. *Parameter estimation*: seek for

$$\underline{\theta}^* = arg \min_{\underline{\theta}} ||\underline{y} - \underline{\tilde{y}}|| \tag{3}$$

where $|| \cdot ||$ is a proper norm, $\underline{y}$ and $\underline{\tilde{y}}$ are N-dimensional vectors of the measured data and of the computed values according to the model $\tilde{f}(\underline{x}, \underline{\theta})$, respectively.

Data-driven approaches

Recently, data-driven approaches to fuzzy modeling have received more and more attention [Abe and Lan, 1995, Horikawa et al., 1992, Jang, 1993, Pomares et al., 2002, Takagi and Sugeno, 1985, Wang and Mendel, 1992, Wang, 1994]. These approaches mainly differ from each other in the way they perform parameter initialization and rule base generation. But, to define the structure of $\tilde{f}(\cdot)$, all of them follow, in outline, the flow given in Fig. 1: structure identification and parameter estimation are mutually related, and are performed within the same loop. The overall procedure loops on increasing model complexity till the obtained model meets a given criterion, such as a prespecified target accuracy or a model evaluation index. The initial model complexity may fix either the number L of partitions of each variable domain, often

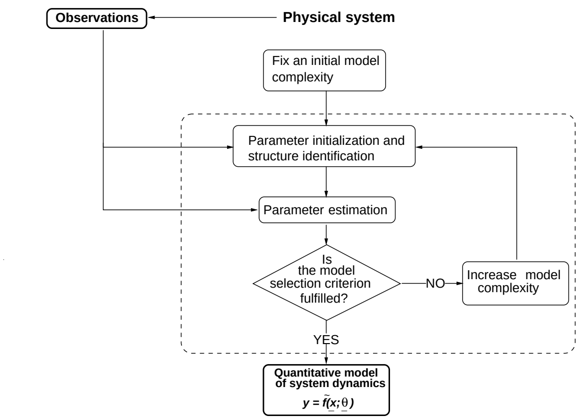

Figure 1: Main steps in data-driven approaches

performed in accordance with clustering techniques [Bezdek, 1981, Sugeno and Yasukawa, 1993], or the number M of rules. In the former case, the domain is splitted into L regions, to which a μ is assigned; in the latter one, the μ's are initialized around M data by directly locating their centers on the data themselves. The rules are generated by determining either all the possible combinations of the μ's [Jang, 1993] or the combinations of those μ's that identify regions where the data pairs get the maximum degrees [Wang, 1994]. Given a fuzzy structure, the parameter vector is optimized through a nonlinear estimation procedure.

FS-QM: a QR-driven approach

The method we propose, sketched in Fig. 2, clearly separates the structure identification phase from the parameter optimization one. FS-QM strongly exploits a QR paradigm, namely QSIM, to drive almost all modeling phases. It builds the rules the model equation (2) is grounded on by encoding the qualitative value and state descriptions of the system dynamics, inferred by the simulation of a QSIM model, into the fuzzy formalism. A crucial issue deals with the proper representation of landmarks and intervals between them in the fuzzy framework since it determines the initial value $\underline{\theta}_0$ of $\underline{\theta}$.

Variable fuzzy partition. Given a generic input/output variable x, and its quantity space Q_x, let us call *qualitative quantity-space* the finite totally ordered set Q_{L_x}, whose elements are landmark values, l_k, and open intervals (l_k, l_{k+1}), bounded by two adjacent landmark values. Let us call *fuzzy quantity-space* the finite ordered set Q_{F_x}, whose elements define the fuzzy partition of the domain of x. Let us assume that quantitative knowledge on the real interval $[a_k, b_k], a_k \neq b_k$, which the landmark l_k belongs to, is given: Q_{Fx} is defined as image of a bijective mapping ν of Q_{L_x}. More precisely:

- $\nu(l_k) = F_{2k-1}$, characterized by $\mu_{F_{2k-1}}(x)$ with support (a_k, b_k), and $\mu_{F_{2k-1}}(x) = 1$ in $x = c_k = (a_k + b_k)/2$;
- $\nu((l_k, l_{k+1})) = F_{2k}$, characterized by $\mu_{F_{2k}}(x)$ with support (c_k, c_{k+1}), and $\mu_{F_{2k}}(x) = 1$ in $[b_k, a_{k+1}]$.

Q_{F_x} is associated with a parameter vector $\underline{\hat{\theta}}_0^{(l)} \in \mathbb{R}^{n_l}$, whose elements are the locations $\{a_k, c_k, b_k\}$ associated with all μ's in Q_{F_x}. The fuzzy partition Q_{F_x} is built for each variable. Then, the system parameter vector $\underline{\theta}_0$ is made up of $(n+1)$ vectors, i.e. $\underline{\theta}_0 = (\underline{\hat{\theta}}_0^{(1)}, \ldots, \underline{\hat{\theta}}_0^{(n+1)})$ where n vectors are

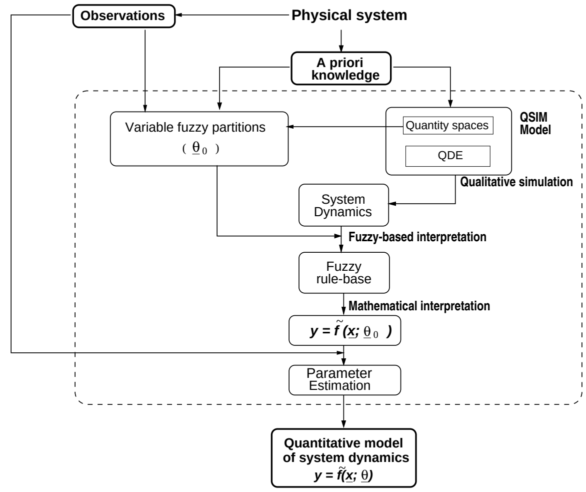

Figure 2: Main steps in FS-QM

defined as above, and $\underline{\hat{\theta}}_0^{(n+1)}$ is made up of the centers of the μ's of the output variable y. Figure 3 exemplifies how $Q_{L_x} = \{l_1, (l_1, l_2), l_2, (l_2, l_3), l_3\}$ is mapped into Q_{F_x}.

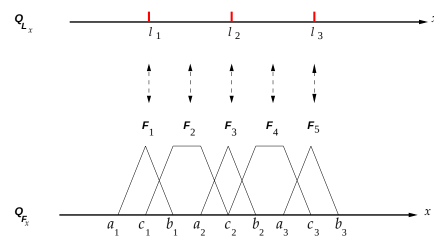

Figure 3: Mapping Q_{L_x} into Q_{F_x}

The mapping ν defines a complete and consistent fuzzy partition, and states a correspondence one-to-one between the landmark-based and the fuzzy-based representation of real values. By definition, the μ's have bounded supports: in this implementation, we have respectively chosen triangular and trapezoidal μ's to represent landmarks and intervals between them. But, due to the universal approximation theorem which holds for the considered class of FS's, other shapes could be chosen to represent the μ's without affecting the approximation capabilities of the resulting fuzzy model.

Fuzzy rule generation. On the basis of the mapping ν, we can automatically translate the finite set of qualitative behaviors $\{B_1, \ldots, B_m\}$ generated by QSIM into fuzzy rules. To optimize the rule number, the behavior tree is conveniently analyzed and preprocessed, and only behaviors representative of significant distinctions are translated into rules. In outline the algorithm, given in [Bellazzi et al., 2001, Guglielmann and Ironi, 2002], maps each admissible behavior into a set of rules. In each rule, the antecedents and the consequent are the fuzzy representation of the qualitative value of all x_i at the current time, and of y at the next time, respectively. Then, each rule gives a measure of the possible transition from one state to the next one. In this way, the entire range of possible system dynamics is embedded into the rule base. As it may happen that identical and/or conflicting rules are generated, filtering procedures are applied, and the final rule base results to be complete and consistent.

Robustness of fuzzy models

By model robustness we mean both generalization capability and numerical stability of the identified model.
Generalization deals with the capacity of the model of computing, for inputs never used to identify the model, an output that correctly reproduces test data drawn in the same experimental conditions as those of the identification data set. The equation structure strongly influences model generalization. But, the model is fully defined after its parameters have been identified from the data samples. Therefore, also the size and the representativeness of the data set are influencing factors of such a property. Under the assumption that the data set is fixed, the goal of achieving good generalization may be pursued with the determination of the optimal model structure.

As for models represented by the equation (2), parameter estimation (3) is generally equivalent to a nonlinear least squares problem. Let us observe that such a problem is ill-posed [Engl et al., 1996] in the sense that solutions do not necessarily depend on the data in a continuous way, or equivalently that solutions are not numerically stable. To obtain a stable solution, we have to use regularization techniques. Whereas the regularization theory is well developed for linear ill-posed problems [Tikhonov and Arsenin, 1977], the nonlinear case is not.

Data-driven approaches

Although they have been shown to be successful in many applications, data-driven approaches still suffer from many drawbacks as their performance strictly depends both on the number and on the quality of data. As the experimental data may be scattered and noisy, the completeness of the FRB might fail, and the model built upon it, although optimal with respect to the fixed criterion, might not be able to capture the true underlying system dynamics. Moreover, if the input dimension or the number of rules are increased, the model complexity grows exponentially, and parameter estimation may become numerically untractable. This could lead to a sub-optimal model structure. Last but not least, if the parameter number of the built model is higher than the number of samples, overfitting phenomena may occur, i.e. the model may also learn the noise present in the data themselves with a consequent loss of the ability to generalize [Bellazzi et al., 2001].

A great deal of work related to fuzzy modeling provides methods for parameter optimization. Since a FS may be represented as a three-layer neural network, most of the approaches are based on variants of the back-propagation algorithm. As a matter of fact, these methods find a local minimum that converges to the optimal value $\underline{\theta}^*$ when a "good" initialization $\underline{\theta}_0$ is given. Then, the definition of "good" partitions of input-output variables is of great importance to a reasonable solution of problem (3). However, even in case data are not corrupted by noise, and their number is large enough, the parameter estimation problem is ill-posed, and some prior knowledge must be exploited to constrain the model search space, or equivalently, problem (3) must be regularized. As for fuzzy modeling, only recently the regularization problem has been addressed [Burger et al., 2002, Johansen and Babuška, 2003] for a particular class of FS's.

QR-based approach: FS-QM

The model built in the FS-QM framework does not suffer from the drawbacks typical of data-driven approaches. The rule base is generated from the QSIM model only, and then scattered and noisy data do not hamper its completeness. Owing to the way it has been generated, it includes all of the possible state transitions. Then, we can surely assert that it is not sub-optimal. However, it may be redundant as it may include rules that represent spurious transitions. Spurious rules are never instantiated by the experimental data. Thus, they influence neither the approximation nor the generalization capabilities of the resulting model, although they may reduce the computational efficiency. The number of parameters, independent of the number of rules and initialized on the basis of prior knowledge, grows linearly with the number of qualitative values, or equivalently with the number of variable partitions. This is an important feature of FS-QM as, rule number being equal, it builds a model with a significantly smaller number of parameters than data-driven approaches. This together with a good initialization of both structure and parameter vector results in its outperformance as for computational efficiency. If the QSIM model does represent the underlying physical knowledge of the system at hand, the resulting equation structure holds good generalization capabilities. This is still valid also when a small number of samples is available: as the model structure is not learned from data, overfitting phenomena do not occur as demonstrated in [Bellazzi et al., 2001]. Let us remark that in the same paper, to make the comparison with a number of different data-driven approaches fair, we considered gaussian μ's, usually preferred in those approaches as their unbounded supports ensure completeness of fuzzy partitions, and then better approximation properties. But, in FS-QM, completeness of both fuzzy partitions and rules is always ensured in itself due to the way they have been built. Figures 4A and 4B illustrate that the generalization capability of FS-QM is preserved independently wheter the chosen functions to represent the μ's have bounded or unbounded supports. More precisely, they show the results obtained by validating models, that consider gaussian/double gaussian μ's and triangular/trapeziodal μ's, respectively, on a data set different from that used to identify the models but gathered in the same experimental conditions.

Although the model equation and the initial guess $\underline{\theta}_0$ have been built on structural knowledge, parameter estimation from data remains an ill-posed problem, and numerical instability may occur unless we further restrict the model search space. This can be done by imposing prior knowledge on the solution, namely by constraining either the function $f(\cdot)$ or the parameter vector $\underline{\theta}^*$ to belong, respectively, to a specific functional space or a specific trust region. Under the assumption that the prior knowledge used to define the initial estimate $\underline{\theta}_0$ is correct, the optimal parameter vector $\underline{\theta}^*$ should be "close" to $\underline{\theta}_0$. Then, we remove ill–posedness by constraining $\underline{\theta}^*$ to be in the neighborhood of $\underline{\theta}_0$. In practice, for each input-output variable we constrain its associated parameter vector $\underline{\hat{\theta}}^{(l)}$ to belong to a "sufficiently small" region centred on $\underline{\hat{\theta}}_0^{(l)}$. The width of such a region strictly depends on the confidence of the knowledge about landmark values: the more precise the available knowledge on the initial locations of the μ's is, the smaller the region defined in the constraint is chosen. The optimization problem (3) is reformulated as a nonlinearly constrained minimization problem as follows:

$$\underline{\theta}^* = arg \min_{\underline{\theta} \in \mathcal{R}} ||\underline{y} - \underline{\tilde{y}}|| \qquad (4)$$

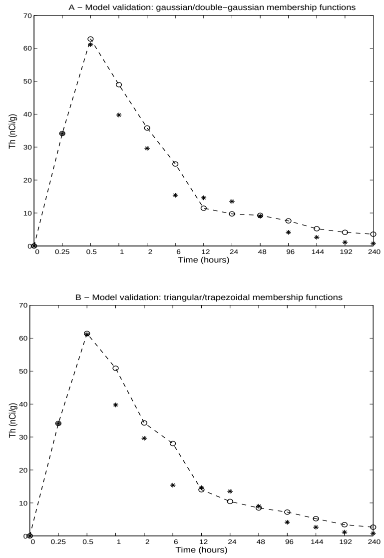

Figure 4: FS-QM validation results: A - Gaussian/Double Gaussian μ's; B - Triangular/Trapezoidal μ's. Data are denoted by $*$, and predicted values by ($\circ$)

where $\mathcal{R} = \{\underline{\hat{\theta}}^{(i)} \in \mathbb{R}^{n_i}, ||\underline{\hat{\theta}}^{(i)} - \underline{\hat{\theta}}_0^{(i)}|| \leq \delta_i\}$; the δ_i's give a measure of the degree of confidence in the initial values of the parameters, i.e. in the prior knowledge. Problem (4) is solved by means of Sequential Quadratic Programming [Nocedal and Wright, 1999], a classical optimization algorithm.

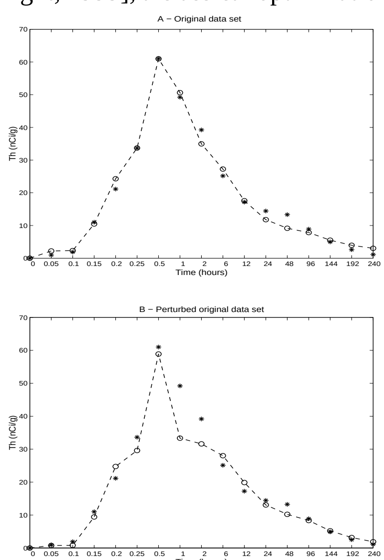

Figure 5: Model identification with unconstrained parameter optimization: A - Original data; B - Original perturbed data.

The solution of the constrained problem is actually made stable, in the sense that "small" perturbations on data do not affect significantly the approximation properties of the identified model, and the estimated values of the parameters. To support this, we show the identification results obtained with two data sets, the second of which is obtained from a perturbation of the first one, in the case of both unconstrained (Fig. 5) and constrained (Fig. 6) parameter optimization. The original data have been perturbed with zero mean normally distributed random noise that is really a small quantity, more precisely its order of magnitude is 10^{-7}. The figures clearly

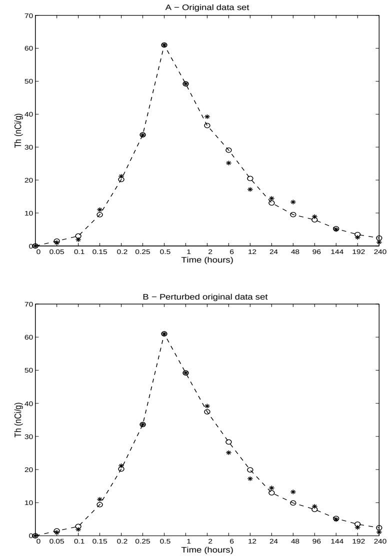

Figure 6: Model identification with constrained parameter optimization: A - Original data; B - Original perturbed data.

highlight that the estimation procedure is sensitive to numerical perturbations, and that it is made actually stable by constraining the minimum problem. The regularized problem is not only more stable but also favors solutions that are easily interpretable.

Interpretability of fuzzy models

Besides meeting robustness requirements, to be really useful, models of dynamical real-world systems should provide a transparent, and intelligible description of their complex dynamics, and parameter values with a clear physical meaning.

The main reason that makes fuzzy systems preferable to other input-output modeling schemes is their capability to express prior knowledge. But, in practice, this feature is not entirely exploited by traditional approaches: they are not able to express deep prior knowledge but heuristics established on the basis of experience and intuition of the domain expert. Then, the resulting model may fail in clearly representing (i) all the information about the system dynamics, and (ii) the physical parameters associated with important changes in the system state. Not to mention data-driven approaches: the definition of a model with good approximation performance very often occurs at the cost of lack of model interpretability.

Factors that influence interpretability are related to both the fuzzy partition and the rule base. On the one hand, the variable partitions should be complete and distinguishable so that each of its subsets may be associated with a clear physical meaning. Distinguishability means that the elements in the partition are separable and strongly consistent. On the other hand, the rule base should be consistent, and made up of a reasonable number of rules: either contradictory rules or their combinatorial explosion, due to a too high dimension of input vector, makes the model very hard to be understood.

To obtain an interpretable model two requirements have to be met: (1) the initial model, namely initial fuzzy partition and rule base, must be interpretable; (2) the model must remain interpretable after parameter estimation.

Data-driven approaches

The drawbacks, mentioned in the previous section, that data-driven approaches suffer from clearly rebound on interpretability [Jin, 2000]. These approaches may lead to incomplete partitions, inconsistent rules, and to an exponential growth of the number of rules and parameters. But, even when these phenomena are suitably controlled, and consequently the initial model is interpretable, such a model feature may vanish after parameter adjustment. During the learning process, the parameters of the μ's may be adjusted so drastically that the resulting fuzzy partition is not complete and distinguishable any more. Regularization techniques may help also to reduce this phenomenon, and then to improve significantly model interpretability [Burger et al., 2002].

Let us underline that, from the physical point of view, the interpretability potential of traditional approaches is, in general, rather weak, even when the identified model meets the conditions for it. As a matter of fact, the model parameters identify regions that do not necessarily correspond to descriptions of the system states physically significant.

QR-based approach: FS-QM

For the way fuzzy partitions are defined by the mapping ν, i.e. (i) complete covering of the variable domain suggested by the knowledge about qualitative values and not by the data, and (ii) μ's with bounded supports, the above mentioned conditions for interpretability, namely partition completeness and distinguishability, are guaranteed. Moreover, a sound physical meaning is associated with each fuzzy set and its parameters as the fuzzy partition of each variable domain is landmark–based. The rule base results to be complete, because it embeds all the significant features of the system dynamics captured by the qualitative behaviors, and fully intelligible, as each rule expresses the transition of the system from a state to its successor. The rules are also consistent: in fact, possible conflicts between them, namely rules with the same IF–part and different consequent, are solved on the basis of the degree of each rule calculated on the data samples. Finally, let us remind that the generation of rule base is grounded on the variable state transitions that occur in the admissible behavior set, i.e., a physically significant subset of the simulated behavior tree. Then, the number of rules is linear with the number of such variable state transitions.

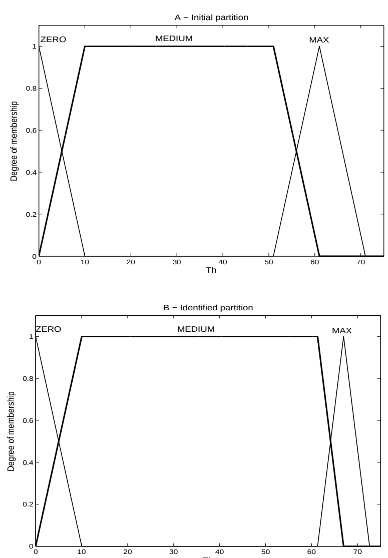

Figure 7: Triangular/trapezoidal μ's: A - Initial; B - Identified.

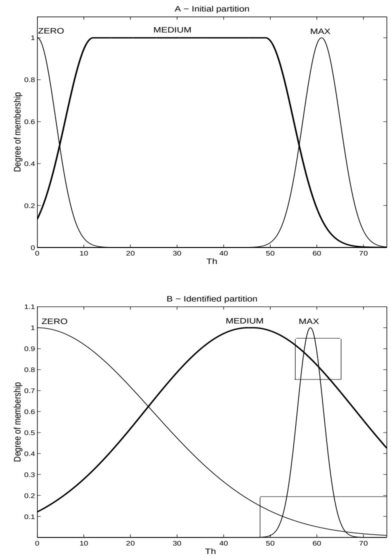

Figure 8: Gaussian/double-gaussian μ's: A - Initial; B - Identified.

Since the model structure is properly initialized on deep knowledge, the parameter estimation procedure refines the numerical ranges to which landmark values are initially bounded. Interpretability of fuzzy partitions is preserved after parameter optimization thanks to the constraint defined in (4): as a matter of fact, such a constraint is able to keep the physical meaning of the parameters, and the consistency and separability properties ensured by the mapping ν (Fig. 7).

Let us emphasize that the hypothesis made on the supports of the μ's, namely that they must be bounded, is necessary to retain the consistency property, and consequently to gain interpretability. As a matter of fact, if the supports were unbounded, such as in the case of gaussian functions, the weak consistency of initial fuzzy partition [Zeng and Singh, 1996] can be lost after the optimization procedure. Figure 8-A shows the initial partition of the same variable in Fig. 7 where gaussian functions are exploited to represent the knowledge about its landmarks. From Fig. 8-B it is evident that interpretability is definitely lost: as highlighted by the regions in the boxes, both weak consistency and separability are not preserved after the optimization procedure.

The model interpretability in "actual" physical terms, made possible by FS-QM, lays the groundwork for new application perspectives of fuzzy models. Let the initial model structure describing the system under normal conditions be the *nominal model*. We could exploit it in a diagnostic context (i) to evaluate the deviations from the nominal values of the parameters identified on new data samples, and then to infer the most plausible diagnostic class; (ii) to test different hypotheses about the underlying physical mechanisms by modifying the qualitative model, and then by simulating it to generate a rule base which better describes the newly observed system dynamics. The first task can be performed thanks to the clear physical meaning of the parameters of the μ's, whereas hypotheses testing is more concerned with the

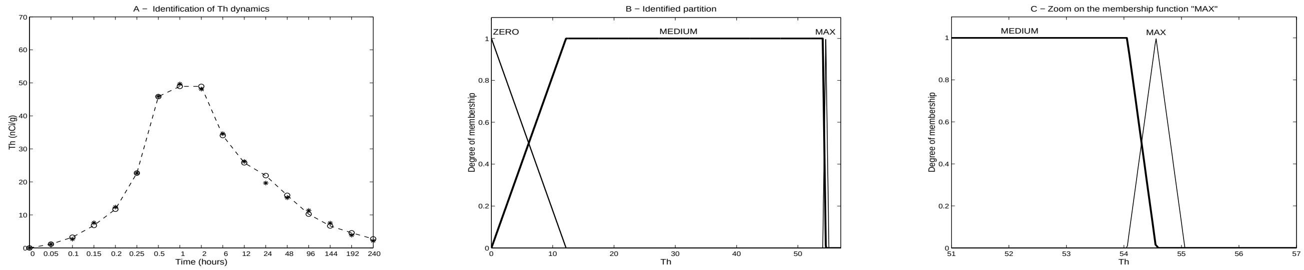

Figure 9: Thiamine dynamics in treated diabetic subjects: A - Identification results; B - Identified partition; C - Zoom of identified partition.

rule base generation. In outline, the diagnostic classification problem may be tackled as follows. When a new data set is available, parameter estimation is performed under the same conditions (same initial μ's and width of the region $\mathcal{R}$ defined in (4)) used to identify the nominal model. If the optimization procedure succeeds in identifying an accurate model where the identified centers of the μ's fall into their initial supports, we can infer that the data set is related to a normal situation. Otherwise, we can slightly relax the constraints on the parameters: if the optimization procedure succeeds, and again the centers of the identified μ's lie in their initial supports, we can conclude that the data set refers to "quasi–normal" conditions. A further unsuccessful result means that the data set actually refers to a system corrupted by a fault or a disease, and prior knowledge on landmark values has to be updated accordingly, i.e. the initial fuzzy partitions have to be changed. As exemplification, let us consider as nominal model that one related to Thiamine kinetics in normal patients [Bellazzi et al., 2001]. We are given three different sets of data measured in as many different experimental settings. The first one, related to Thiamine kinetics in normal patients, has been already used in the paper to refine the parameter values of the nominal model (Fig. 7); the second and the third ones refer to insulin-treated and untreated diabetic subjects, respectively. Let us observe that the centers of the identified μ's in Fig. 7-B do fall into their initial supports, as we expected. As for the second set of data, the optimization procedure fails to identify an accurate model under the same initial conditions, unless we relax the constraints on the parameters, i.e. we enlarge the region $\mathcal{R}$. The results obtained are really satisfactory in terms of approximation accuracy (Fig. 9-A). The optimized μ's in Fig. 9-B mainly differ from those in Fig. 7-B as for the fuzzy set labelled "MAX" zoomed in Fig. 9-C: we can observe that, even if the center moves away from its nominal value, it is still within the support of the initial μ_{max} (Fig. 7-A). Also with the third set of data, FS-QM fails to approximate the system dynamics under the same initial conditions, and a significant enlargement of the region $\mathcal{R}$ reveals to be unsuccessful. Only a drastic change of the initial partitions (Fig. 10) allows us to get good results. Actually, this is due to the inadequacy of the prior knowledge on landmark values related to the physiological system to represent properly the pathological situation.

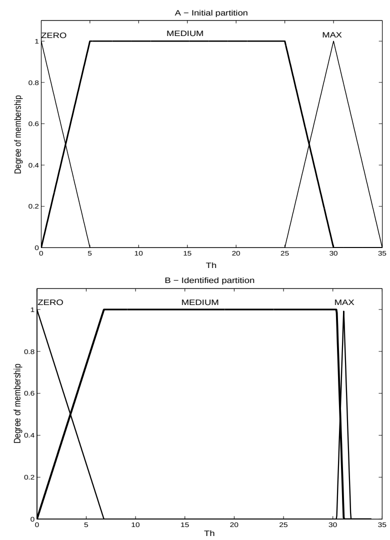

Figure 10: Untreated diabetic subjects: Thiamine fuzzy partition. A - Initial; B - Identified.

Conclusion

Fuzzy modeling aims at achieving an approximation of the unknown functional relation between the input and output variables of a system from experimental data. To achieve a robust and interpretable fuzzy model, FS-QM effectively employs all the available structural prior knowledge, represented in QSIM, and empirical data. The embedment of deep prior knowledge into the FS makes the identification problem better posed, since it properly delimits the model search space. In addition, the prior knowledge allows us to define a good initial estimate $\underline{\theta}_0$, and, then, to define a trust region where $\underline{\theta}^*$ is supposed to belong to. If the prior knowledge is correct, this will lead to a model that has good generalization and interpretability properties also in data-poor contexts. In FS-QM, the gained parameter interpretability from the physical point of view represents an added value that stands chances for fuzzy models to be used to perform a larger spectrum of tasks than the usual one, namely system control. FS-QM models might be conveniently applied, for example, in a diagnostic context. On the one hand, diagnostic hypotheses that explain the observed behaviors could be tested by introducing structural variations into the underlying qualitative model, and by validating the fuzzy model built on the basis of the newly generated rule base. On the other hand, diagnostic hypotheses could be drawn from the analysis of the deviations of the estimated values of parameters from the nominal ones. Future work will thoroughly explore the diagnostic potential of FS-QM models.

A drawback of FS-QM deals with the possible generation of spurious rules. A spurious rule is never instantiated by the data samples, and then its rule degree is equal to zero. Sufficient but not necessary condition for a rule to be sound is that its degree is different from zero. The definition of criteria for either detecting the presence of spurious rules or weighting the "doubtfully sound" rules requires to be investigated to further improve the computational efficiency of FS-QM.

References

Abe, S. and Lan, M. (1995). Fuzzy rules extraction directly from numerical data for function approximation. *IEEE Trans. on SMC*, Vol. 25, No. 1, pp. 119–129.

Bellazzi, R., Guglielmann, R. and Ironi, L. (1999). A qualitative-fuzzy framework for nonlinear black-box system identification, in: T. Dean, (ed.) *IJCAI 99*, Morgan Kaufmann, San Francisco, pp. 1041–1046.

Bellazzi, R., Guglielmann, R. and Ironi, L. (2000). How to improve fuzzy-neural system modeling by means of qualitative simulation. *IEEE Trans. on Neural Network*, Vol. 11, No. 1, pp. 249–253.

Bellazzi, R., Guglielmann, R., Ironi, L. and Patrini, C. (2001). A hybrid input-output approach to model metabolic systems: An application to intracellular thiamine kinetics. *Journal of Biomedical Informatics*, Vol. 34, pp. 221–248.

Bellazzi, R., Ironi, L., Guglielmann, R. and Stefanelli, M. (1998). Qualitative models and fuzzy systems: an integrated approach for learning from data. *Artificial Intelligence in Medicine*, Vol. 14, pp. 5–28.

Bezdek, J. (1981). *Pattern recognition with fuzzy objective function algoritms*, Plenum, New York.

Bradley, E., Easley, M. and Stolle, R. (2001). Reasoning about nonlinear system identification. *Artificial Intelligence*, Vol. 133, No. 1-2, pp. 139–188.

Burger, M., Haslinger, J., Bodenhofer, U., and Engl, H. (2002). Regularized data-driven construction of fuzzy controllers. *Journal of Inverse and Ill-Posed Problems*, 10(4):319–344.

Capelo, A., Ironi, L. and Tentoni, S. (1998). Automated mathematical modeling from experimental data: an application to material science. *IEEE Trans. on SMC*, Vol. 28, No. 3, pp. 356–370.

Engl, H. W., Hanke, M. and Neubauer, A. (1996). *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht.

Girosi, F., Jones, M. and Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation*, Vol. 7, pp. 219–269.

Guglielmann, R. and Ironi, L. (2002). Strategies for improving neuro-fuzzy identification of nonlinear dynamical systems, in: *Proc. FSKD'02*, Singapore, Vol. 1, pp. 59–63.

Haykin, S. (1997). *Neural networks - a comprehensive foundation*, Prentice Hall, New York.

Horikawa, S., Furuhashi, T. and Uchikawa, Y. (1992). On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm. *IEEE Trans. on Neural Networks*, Vol. 3, No. 5, pp. 801–814.

Jang, J. (1993). Anfis: Adaptive network based fuzzy inference system. *IEEE Trans. on SMC*, Vol. 23, pp. 665–685.

Jin, Y. (2000). Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement. *IEEE Trans. on Fuzzy Systems*, Vol. 8, No. 2, pp. 212–221.

Johansen, T. (1994). Fuzzy model based control: stability, robustness, and performance issues. *IEEE Trans. on Fuzzy Systems*, Vol. 2, No. 3, pp. 221–234.

Johansen, T. and Babuška, R. (2003). Multi-objective identification of Takagi-Sugeno fuzzy models. *IEEE Trans. on Fuzzy Systems*, Vol. 11, No. 8, pp. 847–860.

Kay, H., Rinner, B. and Kuipers, B. J. (2000). Semiquantitative system identification. *Artificial Intelligence*, Vol. 119, No. 1, pp. 103–140.

Kuipers, B. J. (1994). *Qualitative Reasoning: modeling and simulation with incomplete knowledge*, MIT Press, Cambridge MA.

Ljung, L. (1987). *System Identification - Theory for the User*, Prentice-Hall, Englewood Cliffs.

Mamdani, E. (1974). Applications of fuzzy algorithms for simple dynamic plant. *Proc. IEE*, Vol. 121, No. 12, pp. 1585–1588.

Niyogi, P. and Girosi, F. (1999). Generalization bounds for function approximation from scattered noisy data. *Advances in Computational Mathematics*, Vol. 10, pp. 51–80.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*, Springer, New York.

Poggio, T. and Girosi, F. (1990). Networks for approximating and learning. *Proc. IEEE*, Vol. 78, No. 9, pp. 1481–1497.

Poggio, T. and Smale, S. (2003). The mathematics of learning: dealing with data. *Notices of the American Mathematical Society*, Vol. 50, No. 5, pp. 537–544.

Pomares, H., Rojas, I., Gonzales, J. and Prieto, A. (2002). Structure identification in complete rule-based fuzzy systems. *IEEE Trans. on Fuzzy Systems*, Vol. 10, No. 3, pp. 349–359.

Sugeno, M. and Yasukawa, T. (1993). A fuzzy-logic based approach to qualitative modeling. *IEEE Trans. on Fuzzy Systems*, Vol. 1, No. 1, pp. 7–31.

Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. on SMC*, Vol. 15, No. 1, pp. 116–132.

Tikhonov, A. and Arsenin, V. (1977). *Solutions of ill-posed problems*, Winston, Washington DC.

Wahba, G. (1990). *Splines Models for observational data*, Vol. 59, Series in Applied Math., SIAM, Philadelphia.

Wang, L. (1994). *Adaptive Fuzzy Systems and Control: design and stability analysis*, Englewood Cliff, NJ:Prentice–Hall, University of California at Berkeley.

Wang, L. and Mendel, J. (1992). Generating fuzzy rules by learning from examples. *IEEE Trans. on SMC*, Vol. 22, No. 6, pp. 1414–1427.

Zeng, X. and Singh, M. (1996). Approximation accuracy analysis of fuzzy systems as functions approximators. *IEEE Trans. on Fuzzy Systems*, Vol. 4, No. 1, pp. 44–63.