Fuzzy Automata Identification Based on Knowledge Discovery in Datasets for Supervision of a WWT Process

Julio Waissman *, Cherif Ben-Youssef * and Gabriela Vázquez **

*Centro de Investigación en Tecnologías de Información y Sistemas, Universidad Autónoma del Estado de Hidalgo, Carr. Pachuca-Tulancingo, Pachuca, Hidalgo, Mexico.

julio@uaeh.reduaeh.mx
cherif@uaeh.reduaeh.mx

**Centro de Investigaciones Químicas, same address.

gvazquez@uaeh.reduaeh.mx

Abstract: This paper describes a methodology for the design of a supervisory system applied to a wastewater treatment process. A behavioral model is build by the joint participation of the process expert together with clustering techniques applied to measured signals. A fuzzy automaton becomes a heuristic model of the process under supervision. In the application examples, a real data for an activated sludge wastewater plant was used. The automaton states are identified by experts as the significant operation situations and the measured variables of the plant generate the transitions. Thus, this model reflects the dynamics of the continuous system underneath.

Keywords: Process Supervision, Fuzzy Automata, Data Analysis, Application.

1 Introduction

The management, control and supervision of a wastewater treatment process (WWTP) are complex and non trivial tasks, mainly due to the features of the process, and to the catastrophic consequences that could be achieved by an incorrect operation. Furthermore, the necessity to reduce the costs of process supervision and to prevent damage has yielded to a great number of monitoring and supervision applications in the industrial area. The primary objective of WWTP is to reach the specified requirements on the outflow water quality, in order to restore the natural environmental balance broken by human activity. The WWTP are complex MIMO systems which exhibit a wide range of dynamic behavior: the inflow is variable in quality and quantity, and microbial population also changes over time in quantity and in the relative number of species. In another hand, the measured variables are insufficient; there exists a few and unreliable on line analyzers. Most of the reliable data measured is from the environmental (measured or controlled) variables, such as temperature, pH, DOT, etc. These reasons make the use of traditional techniques difficult or even impossible.

Despite these problems, operators of WWTP are capable of identifying normal and abnormal situations, and consequently, of taking the suitable actions. Process experts are able to draw such conclusions by analyzing a set of measured signals collected from the plant. This suggests that, despite the lack of process models, measured data can be used instead in the supervisory system development. In a supervisor system two main tasks are to be considered: process monitoring and supervisory control (Kotch, 1993). Process monitoring includes data collection and processing in order to have an updated plant state knowledge. At this stage the system must be able to decide whether the process is in an abnormal state and whether a corrective action should be undertaken. Such intervention along with diagnosis is regarded under the supervisory control task.

In this paper, a methodology is proposed to build a fuzzy automaton as a reliable model for process monitoring and supervisory control design. The
complexity of the system imposes the co-operation of the two different kinds of information available: the raw data from past operations and the expert knowledge. This is obtained applying the LAMDA classification technique of data, then, a finite set of clusters is build and identified by the expert to real situations.

2 The fuzzy automata

For supervision purposes, they have been a tradeoff between model complexity and our ability to perform analysis on the systems via the model. Modelling techniques for supervision must be able to support a macroscopic view of the dynamical system. The discrete event system (DES) models are a well suited approach to perform a supervision task. This kind of models ignores intentionally some of the system characteristics, specifically those that need not to be considered in attempting to meet a particular performance specification.

However, the biological phenomena in a WWTP do not lend themselves to crisp transitions, and the adequation of the data samples into the states is often a matter of degree, rather than a yes-or-no decision. A fuzzy generalization of a DES model, like finite automata, is better suited for this kind of systems. A fuzzy automaton is one which has a possibility measure associated with it. Intuitively, there are two ways of associating a possibility measure with automata. One is to associate a fuzzy membership function to transitions. We can talk about a Non-deterministic Fuzzy Automaton, since the next state is not fully determined by the current state and the current input event. This kind of fuzzy automata has a behavior similar to a Markov model.

The second way is to take a deterministic finite automaton and associate a possibility function to each state; thus, in each state, the adequation to keep the same state or to change to an adjacent state is computed. This is a Deterministic Fuzzy Automaton, because, once the decision was taken, the next state is fully determined. We will use this approach in order to take advantage as much as possible of the developments in the analysis and control design of deterministic finite automata for a future supervisor design. This includes results on controllability, observability, stability and algorithms for control design among others (Ramadge & al., 1989).

In this paper, in order to cope with numerical values from the measured variables of the plant, we use a slightly different definition of fuzzy deterministic automata that the one proposed in (Chan, 1971). A Fuzzy Deterministic Finite Automaton is defined as a quintuple $F = (\Sigma, Q, q_0, f, \lambda)$, where $\Sigma \subseteq \mathbb{R}^n$ is the vector of the $n$ input values, $Q$ is the set of states, $q_0$ the initial state, $f : Q \times \Sigma \rightarrow Q$ is the feasibility function and $\lambda : Q \times \Sigma \times Q \rightarrow 0$ is the performance function. Both functions are allowed to be partially defined.

The feasibility function $f$ gives the objective constraints for the automata to change from one state to another. For example, if $f(q_i, x) = q_j$ then when the automata is in state $q_i$ and the input vector is $x$, the actual state changes from $q_i$ to $q_j$. On the other hand, the performance function $\lambda$ gives a subjective performance evaluation for the automata to change from one state to another in a particular situation. For example, if $\lambda(q_i, x, q_j) = 0.9$ and $\lambda(q_i, x, q_k) = 0.5$, then when the automata is in situation $q_i$ and the input vector $x$ is received, subjectively the transition from $q_i$ to $q_j$ is more adequate than the transition from $q_i$ to $q_k$.

Intuitively, in a maximum likelihood approach, $f$ can be defined by:

$$f(q_i, x) = \max_{q_j} \{ \lambda(q_i, x, q_j) \} \quad (1)$$

In our approach, $\lambda$ is a set of membership functions between adjacent states (including themselves) and all the available measured data of the plant. This function will be determined by fuzzy classification using the LAMDA method.

3 The LAMDA method of classification

The objective of classification is the identification of structures in data. Methods developed in the field of Soft Computing, such as fuzzy logic, are becoming increasingly popular. Such methods offer an attractive alternative to statistical approaches as they do not require a priori assumptions of statistical models.

In (Piera & al., 1989) a fuzzy method of conceptual clustering and classification called LAMDA (Learning Algorithm for Multivariate Data Analysis) was presented. The LAMDA algorithm computes the degree of adequation of an object to a class with all the partial or marginal information available.

The difference between this algorithm and the classical clustering and classification approaches is that LAMDA models the total indistinguishability or homogeneity inside the feature space from which the information is extracted. This is done by means of a special class, called non-informative class (NIC), which accepts all the objects in the same status. Therefore, the adequation degree of these class objects acts as a minimum threshold to assign an element to a significant class. The minimum threshold is therefore not fixed arbitrary, but is automatically determined by the proper context.

Consider a set of objects or situations $X$ and a set of attributes of finite cardinal $n$. An object is represented by a $n$-component vector $x$ where $x_j \in A_j$ is the value taken by the $j^{th}$ descriptor of that object. The attributes can be of qualitative or quantitative type. To make possible a direct confrontation between objects and classes, the latter must be described with
reference to the same attribute used for the observations. Given an object \( x \) and a class \( C_j \in C \), LAMDA computes for every attribute the so called marginal adequacy degree \( \text{MAD}_j \) between the value that the attribute \( j \) takes over \( x \) and the value that the attribute takes over \( C_j \). Thus a \( \text{MAD} \) vector can be associated with object \( x \). This vector has a number of components equal to the number of attributes. \( \text{MAD} \) is a membership function derived from a fuzzy generalization of a binomial probability law:

\[
\text{MAD}_j = \rho_j(x_j, c_j) (1 - \rho_j)^{1 - \nu(x_j, c_j)} \tag{2}
\]

where \( \nu: A_j \times A_j \rightarrow [0, 1] \) is a presence function between the value of the \( j \)th attribute of object \( x \) and a prototype of class \( C_j \) for the attribute \( j \), \( c_j \). So, if \( x_j = c_j \) then \( \nu(x_j, c_j) = 1 \). \( \rho_j \) is the possibility of the object \( x \) to belong to class \( C_j \) concerning only that attribute. Remark that when \( \rho_j = 0.5 \), for every distance function \( \nu \), \( \text{MAD}_j \) is equal to 0.5. Thus, the NIC class is characterized by a \( \text{MAD} \) vector with \( \rho = 0.5 \) for every \( j \).

All previously calculated marginal information given by the \( n \) \( \text{MAD}_j \) is used to compute the global adequacy degree (\( \text{GAD}_i \)) of each class. \( \text{GAD}_i \) is obtained by summarizing the \( \text{MAD}_j \) through a logical aggregation operator \( L \). Two properties required for such operators are commutativity and monotonicity. By commutativity we mean that the order in which we index the \( \text{MAD}_j \) functions does not effect the classification. By monotonicity we mean that as the value of an individual \( \text{MAD}_j \) increases, so does the \( \text{GAD}_i \). To connect all \( \text{MAD}_j \), LAMDA uses a type of aggregation operators called mixed connectives of linear compensation, introduced and studied in (Piera & al., 1991). These operators are located within the framework of fuzzy sets theory. In LAMDA the mixed connectives used are derived from the following linear convex combination:

\[
L_\alpha = \alpha T(\cdot) + (1 - \alpha) N(\cdot) \tag{3}
\]

where \( \alpha \) is a parameter belonging to unit interval, \( T(\cdot) \) is a continuous \( n \)-norm and \( N(\cdot) \) its dual \( n \)-conorm. Data for each class and each attribute are separately used to obtain the corresponding estimate parameters. Parameters \( \rho_j \) and \( c_j \) for each class and for each descriptor are estimated by minimizing a maximum likelihood criterion. Consider a subset of \( X \) of finite cardinal \( t \) where all objects belong to class \( C_i \) being \( x_k \) the value that the \( k \)th object \( x \) of the subset takes in the \( j \)th attribute. The estimated parameters, \( \rho_j \) and \( c_j \), minimize the maximum likelihood criteria if, and only if, the following conditions are satisfied:

\[
\rho_j = \frac{1}{t} \sum_{i=1}^{t} p(x^i_j, c_j) \tag{4}
\]

In (Waissman, 2000), functions to compute the \( \text{MAD}_j \) parameters for three different distance functions \( \nu \) are proposed together with the corresponding sequential learning equations. For a clustering algorithm, each \( x \) is classified with the existing classes. If a NIC is chosen, a new class is created by initializing it with the present \( x \). By this procedure, there is no need of knowing beforehand the number of classes. Because the mixed connectives from equation (3) depend on a parameter \( \alpha \), called exigency index, this leads to the concept of families for these operators. Therefore, given such family, it is possible to associate different classifications with the same data set, depending on the value chosen for \( \alpha \). The method is more exigent as the number of objects assigned to NIC class increases and therefore the number of clusters. In (Piera & al., 1991), the authors have shown that increasing the value of \( \alpha \) also increases the exigency of cluster algorithm. Thus, by changing the value of \( \alpha \), different partitions from the same data, based on the same logical criterion, can be obtained.

4 Identification methodology

When only expert knowledge is used to identify process situations or states, any of these situations can arise: the expert can express only a partial knowledge from process; he does know the existence of several states but he ignores how to recognize them from on-line data; he doesn't have a clear idea on which states to recognize.

For biochemical processes, biotechnologists can apply expert rules when recognizing some of the physiological states from on-line data. Nevertheless those rules usually don’t take into account other phenomena that can change the evolution of signals without any influence in the physiological state. This leads to wrong conclusions. It is mainly due to the fact that the expert is not able to draw conclusions from the analysis of multiple signals between which there exist true relationships. Nevertheless, a classification tool copes well with this drawback. This proves the need of an iterative methodology to identify the biological states, which refines the expert knowledge with the analysis of past data sets.

Knowledge Discovery in Datasets (KDD) techniques is applied to extract information from different past raw data records. The aim of the KDD techniques is to discover the substructures in the feature space of the available data. This is achieved in an unsupervised manner. This results in a partition of the feature space that represents more closely the natural structures in the data. Process states, causal
relationships between them and transition conditions are identified by running LAMDA. However, expert knowledge is required in order to validate those results by constraining them according to the physical laws which rule the real process.

So the proposed methodology uses KDD techniques under the supervision of the expert to obtain a model of the process (Waissman & al., 2003). It is based on the iterative application of LAMDA in unsupervised mode, to identify the set of states with a physical meaning for the expert and it can be expressed as follow: initially, given a set of measured data, no knowledge on process states is take into account (i.e., all data correspond to the same general state) and apply unsupervised learning to data in order to obtain a set of classes. Then, the expert must map the set of classes to a set of physical states. Three situations are possible: a class is equivalent to a biological state; a set of classes is equivalent to a biological state; or a class is not equivalent to any known state. If there exist classes in the last situation, apply clustering considering just the data comprising these classes. When all data is classified in known states, identify all possible state sequencing paths and apply supervised learning to data on transitions in order to obtain its associated performance functions $\lambda$.

The whole methodology is schematized in Fig. 1.

5 WWTP application example

New directives and regulations have guaranteed the appearance of specific plants to treat wastewaters, being activated sludge systems the most extensively used. In an activated sludge process, the wastewater enters an aerated tank where it is mixed with biological floc particles. After enough contact time, this mixture is discharged to a settler that separates the suspended biomass from the treated water. Most of the biomass is recirculated to the aeration tank again, while a little amount is purged daily.

The database used in this work corresponds to a WWTP plant sited at Sallent, in Catalonia, Spain. This database was taken in normal operation of the plant. The database corresponds to three months of operation (March, July an October 1997) with one measurement per day. The available on-line measured variables are:

- inflow rate $Q_i$ (m$^3$/day) (Fig. 2),
- liquor recycle rate (m$^3$/day),
- pH in the tank (Fig. 2),
- dissolved oxygen in inner reactor (ppm),
- dissolved oxygen in the middle reactor (ppm) (Fig. 2), and
- dissolved oxygen in the outer reactor $O_m$ (ppm).

The sampling time of data is bigger than the dynamic transition of the biological system. The biological dynamics are obtained by the reactor volume, inflow and recycle rates and are defined by the time needed to renew completely the microorganism population. Thus, dynamic information is not needed. Therefore, to deal with dynamics, we need to introduce additional variables carrying temporal information. The sliding windows approach developed by (Sarrate & al., 1998) will be followed in order to generate a trend for each variable. This has been proved to be a convenient way to extract information from continuous signals.

The WWTP data (Fig. 2) come from a new plant and then no faulty states was observed. Nevertheless, a set of different normal states can be identified. With this data we identify 4 different states: a normal operation, with a relative low charge (A); a normal operation, with a relative high charge (B); a relatively
low conversion rate (C); and finally a presence of oxygen in the anaerobic phase (D). In a first iteration, the data from class (D) is clearly differentiated from the other data. That is clear, if we take into account that the state is completely correlated with data from dissolved oxygen in the middle reactor. In a second iteration state (C) is manifestly observed. Finally, the states (A) and (B) are well differentiated in a wide range of the exigency index and n-norm used. Next, the direct observation of all state sequencing for the 3 raw data records leads to the automaton structure shown in Fig. 3, which has been validated by the expert. Finally through a supervised classification, the performance functions of the fuzzy automata that determine the automaton transitions are found.

In Fig. 4, a simulation of the identified model is shown. Data from June are used as input. We can observe, in the lower graph, the membership degree of all functional states and, in the upper the graph, the state path followed during simulation. Each state has its own partition of the feature space, and when a state transition is applied, the adequation degree changes suddenly. We can see that membership degree evolution offers an extra information that can be used to observe and estimate the tendency of states transitions.

**Conclusion**

This application example has shown the interaction between the classification tool, LAMDA, and the process expert knowledge in the identification of a fuzzy automaton describing the behavior of a WWTP. The classification algorithms exploit as much information contained in raw data as possible whereas the expert constraints such information according to his knowledge. Thus, the automaton obtained through this methodology is validated by past raw data as well as by the expert and it is readily available to be implemented as a subsystem of a more global supervisory system.

**References**


