Application of Artificial Neuro-Fuzzy Logic Inference System for Predicting the Microbiological Pollution in Fresh Water

S. Bouharati, K. Bennahammed, D. Harzallah and Y.M. El-Assaf
1Laboratory of Intelligent Systems, Department of Electronic, Faculty of Engineering, UFA Setif-Algeria
2Microbiology Laboratory, Department of Biology, Faculty of Science, UFA Setif-Algeria
3Faculty of Engineering, American University of Sharjah, UAE

Abstract: The classical methods for detecting the microbial pollution in water are based on the detection of the coliform bacteria which indicators of contamination. But to check each water supply for these contaminants would be a time-consuming job and a quality operators. In this study, we propose a novel intelligent system which provides a detection of microbiological pollution in fresh water. The proposed system is a hierarchical integration of an Artificial Neuro-Fuzzy Inference System (ANFIS). This method is based on the variations of the physical and chemical parameters occurred during bacteria growth. The instantaneous result obtained by the measurements of the variations of the physical and chemical parameters occurred during bacteria growth-temperature, pH, electrical potential and electrical conductivity of many varieties of water (surface water, well water, drinking water and used water) on the number Escherichia coli in water. The instantaneous result obtained by measurements of the inputs parameters of water from sensors.

Key words: Testing water, E. coli detection, fuzzy expert system, hybrid intelligent system, artificial neuro-fuzzy logic inference

INTRODUCTION

The classical microbiological analysis in water pollution relies highly upon the operator’s experience. The bacteriological quality of water, gagged by the presence of coliform bacteria, is easily tested in the laboratory. But in medical and industrial microbiology is now a need for «Rapid methods» that will shorten the time between receipt of specimen or sample and the issue of a report. The mechanical, electrical or electronic procedures alone or in combination, can be removing the need for direct human action (Collins et al., 1995). Conductance or impedance measurements are used e.g., for rapid measurement of total microbial activity in many products. An alternative to the traditional direct conductimetry has become available, namely the indirect conductimetry in which the evolution of pH and electrical potential from culture media as a result of growth of microorganisms can be studied (Owens et al., 1989). Some microbiological tests based on the indirect method by the measures of the physical and chemical variations during bacteria growth in water. The level of bacteria control is regarded as one of the most important factors in environmental protection and in hygiene public health. In this study, modelling methods are used for predicting behaviour of a complex system, such as Escherichia coli bacteria which indicators of contamination concentration in water, which have a high non-linearity and a complex structure. The data used in this study were Temperature (T), pH of the water, Electrical Potential (EP) and Electrical Conductivity (EC). The difficulties are embedded in these measurements especially in the values of parameters and their origins. The Artificial Neuro-Fuzzy Inference Systems (ANFIS) can be used to model nonlinear relationships. ANFIS have been employed to learn numerical data recorded from sensory measurement. After being trained, ANFIS store knowledge in numerical weights and biases that are often regarded as a black box scheme. ANNs can be translated to symbolic/linguistic rules extracted are then used as a knowledge base for a fuzzy expert system. (Fharyung and Gary, 2003). Hybrid intelligent architecture tends to be more appropriate in applications that require both numerical computation for higher generalization and symbolic/linguistic reasoning for explanation, it is found that hybridization between symbolic/linguistic and numerical representation can achieve higher accuracy compared to either one alone (Tan, 1997, Taha and Ghosh, 1999, Wermier and Sun, 2000).

Neuro-fuzzy system: One of the major challenges of microbiological detection in drinking water is the

Corresponding Author: S. Bouharati, Microbiology Laboratory, Department of Biology, Faculty of Science, Setif University, Setif 1900, Algeria

309
detecting of bacteria in an unstructured environment, in particular where the physical and chemical properties of the bacteria are not known a priori. The resultant uncertainty makes it difficult to detect nature. Various techniques have been applied to solve this problem. Some approaches are analytic and cannot be easily implemented in real-time applications. Also, the analytic approach cannot be used if variables such as the physical's weight are unknown. To overcome this situation, other approaches using fuzzy detectors have been developed using a number of different sensors to measure the physical variables. The fuzzy control can be used in conjunction with powerful automatic learning methods (i.e., neural networks) as a neuro-fuzzy system (Brown and Harris, 1994; Kecman, 2001a).

In this study, we explore these issues using a very simple four-fingered gripper as an experimental system.

**MATERIALS AND METHODS**

**Fuzzy logic modeling:** In traditional set theory, something either belongs to a set or does not depending on whether it fits the definition for that set. In fuzzy set theory, something can partially belong to a set. For example, let us assume, for illustrative purposes only, that the variable 'number of bacteria' has a range of values where safe is considered to be \(<10^4\) mL\(^{-1}\), whereas low is below this range and 'contaminated' above this range. In traditional set theory, a number of bacteria in water of \(10^3\) mL\(^{-1}\) would be classified as safe, whereas \(10^4\) mL\(^{-1}\) would be contaminated (Fig. 1). With fuzzy sets, a value for a variable can partially belong to a set and have a degree of membership anywhere between zero and one (i.e., 0 ≤ μ ≤ 1) (Fig. 2). Relationships among fuzzy sets are expressed as a series of 'if-then' rules to form a rule base. One of the most widely known fuzzy logic modeling algorithms is that of Sugeno-Yasukawa. The following is a general description of the basic steps in the process, using a microbiological modeling example (Beth et al., 2002).

**Assemble input-output dataset:** The first step is to collect all the variables (e.g., origin, temperature, pH, electrical potential and electrical conductivity of the water) and output (e.g., the number of bacteria). In our study, 1000 samples of water are used. The data were studied in Microbiological Laboratory of the Setif University Algeria. They were collected from different varieties of water (well water -250 samples; surface water -250 samples; used water -250 samples and drinking water -250 samples). The physical and chemical changes, pH, Electrical Potential (EP), Electrical Conductivity (EC) were measured at different Temperatures (T) according the number of bacteria (LogN). The bacterial analysis concerned the *E. coli* bacteria (fecal coliforms) measured by the classical methods-NPP in each case (Table 1).

**Membership function:** The membership function for the input variables is shown in Fig. 3. (Temperature as example). The values of k in the figure for all inputs (k = k1, k2, ..., kn) are equal to the mean values of each time series (Chen et al., 2000). Using these mean k values, the data for the inputs were classified into three linguistic categories: Low, middle, high. The data for the output was classified into three linguistic categories: safe, contaminated, more contaminated Fig. 4.

**Cluster the output:** The values of the output (response) are fuzzy clustered using for examples the fuzzy c-means
The physical and chemical changes, pH, Electrical Potential (EP), Electrical Conductivity (EC) were measured at different Temperatures (T) are chosen as input of the system. The number of bacteria (LogN) is chosen as the corresponding output. Both the different levels of input and output are defined by specific membership function for the fuzzy sets. The models have a multivariable system with N input variables and M output variable.

**Fuzzy rules:** The rules determined by the choice of the fuzzy membership function are defined for each input variable. In general form, each fuzzy rule is written as were A_i and A_j are the fuzzy sets that describe the nature of the inputs, such as small, medium, or high.

**Examples 1:** If Origin is (1) and T is low and pH is medium and EP is low and EC is high THAN LogN is safe.

...Also.

The linguistic control rules of this system are given by: (Li-Xin Wang, 1997).

IF X_i is X_i(1) and X_j is X_j(2) and ... X_k is X_k(1) THAN Y_i is Y_i(1).

**Neural network model:** The study consisted of training and testing of the ANN for prediction of outcome in collected water samples. The data on all measured parameters were prospectively collected in a computer database (Microsoft Excel). ANN was applied to provide a nonlinear relationship between inputs variables (Temperature, pH, electrical potential, electrical conductivity and origin of water) and the output variable (LogN).

In present study, a network with one hidden layer was selected. The output layer comprised a single neuron corresponding to the value of the dependent variable to be fit to or predicted the number of bacteria. The model used was standard, three layer, back-propagation, neural network with N input nodes, L hidden nodes and K output nodes, which can be mathematically expressed as:

$$O_{ok} = f_k \left( \sum_{l=1}^{L} w_{kl} f_l \left( \sum_{j=1}^{N} w_{jl} X_j + b_l \right) + b_k \right)$$

∀k ∈ 1, 2, ..., K.

Where, $O_{ok}$ is the output from the kth node of the output layer of the network for the Pth vector (data point), $X_{pi}$ the inputs to the network for the Pth vector (data point), $W_{kl}^p$ the connection weight between the fth node of the hidden layer and the kth node of the output layer.
representation of knowledge and those of neural networks, which deal with implicit knowledge that can be acquired by means of learning (Harris et al., 2002; Brown and Harris, 1994b; Keeman, 2001b). They mimic human decision processes, in that they manage imprecise, partial, vague or imperfect information. Also, they are able to resolve conflicts by collaboration and aggregation. Moreover, they have self-learning, self-organising and self-tuning capabilities, with no requirement for prior knowledge of relationships of data although they can use this if it is available. They have fast computation using fuzzy number operations. As do fuzzy systems, neurofuzzy systems embody rules (defined in a linguistic way) so that system operation can be interpreted in a transparent fashion (Fig. 7).

The number of rules is given by $P = \prod N_i$, where $i$ is the number of inputs and $N_i$ is the number of antecedent fuzzy sets for input $i$ while $M$ is the number of consequent fuzzy sets. In addition, $Q = \sum N_i$ and $R = \sum_{i=0}^{Q-1} N_i$. Figure 7, shows the implementation of a neurofuzzy system as a feed forward network with three layers, namely: the fuzzification layer, the fuzzy rule layer and defuzzification layer. The mapping between the first two

**Neurofuzzy systems:** Neurofuzzy systems combine the advantages of fuzzy systems, such as transparent

![Fig. 5: The topology of a five-input, one-hidden and one-output network](image)

![Fig. 6: The bacteria number variations in different varieties of water. The values of testing steps are totally confused with the training values](image)
layers is nonlinear and linear between the last two layers (Jantzen, 1998). The fuzzification layer consists of two components: the measurements \( x \), which are the input signals from the system under control and the environment and the computation of the antecedent membership function value, \( \mu A_i(x) \) termed the grade of membership. The fuzzy rule layer then calculates the rule firing strength, which indicates how well the conditions in the antecedent are satisfied. An input \( x \) fires a rule to a degree, \( w \in [0,1] \). The calculation of the rule firing strength is performed using the membership function values for fuzzy sets used in the rule antecedent. The product combiner is the fuzzy implication method, used because it preserves the original shape of the fuzzy set and it is differentiable. Finally, the defuzzification layer converts the values
delivered by the fired rules at the previous layer into analogue output values.

The concepts of fuzzy sets and membership functions: By using fuzzy sets, we can formulate fuzzy if-then rules, which are commonly used in our daily expressions. We can use a collection of fuzzy rules to describe a system's behavior; this forms the fuzzy inference system, or fuzzy controller if used in control systems. In particular, we can apply neural networks, learning method in a fuzzy inference system.

In this study, we are going to use first-order Sugeno fuzzy inference system exclusively, where the output equation of each rule is a linear equation.

This slide explains the basic architecture of an ANFIS. We have to graphical representation of the process of fuzzy reasoning. For each operation of fuzzy reasoning, we can put it into a node in an adaptive network. For instance, given two input $x$ and $y$, first we need to find the membership grades; this is represented by the four blocks in the first layer and each of them generates a membership grades. The firing strengths are computed by nodes in the second layer. The results of the variations of bacteria variation in different varieties of water using ANFIS system are shown in Fig. 8.

CONCLUSIONS

The Artificial neural network system shows that the output-number of bacteria in water corresponding to any variation in the input parameters. The result of the ANN program so far, is a numeric and symbolic terms of number of bacteria in water, using the input data from sensors and output data in the universe of discourse (safety, contaminated; more contaminated). if the number of bacteria in output is around 5 ($\log N \approx 5$), the water is contaminated; if it is less than 4, the water is safe; if it is more than 6 it is more contaminated.

In acute microbiological detection in fresh water, there are based on simple and easy-to-handle concepts like laboratory methods. By the classical methods of water contamination detection would be a time-consuming job and a qualify operators. The bacteria growth in water produces many variations in physical parameters. One of the problems in water quality modeling is the vagueness in the values of pollutant sources and of biochemical coefficients, arising either from natural randomness in time and space or from indirect measurements and limited number of samples. In this study, we used different water varieties with different contamination degree measured by the classical methods. The data were analyzed by the neuro-fuzzy logic modeling technique in an attempt to predict the microbiological pollution in fresh water. With the neuro-fuzzy modeling, we can represent imprecise data and produce imprecise output in the form of fuzzy numbers. For any predictive instrument to be useful in making a triage decision, an important feature is that only data are readily available from sensors. From the results obtained by this study, appear to be a useful tool for future water-testing on pathogenic risk identification, quantification and development of early warning systems for fresh water quality. There room for improvement in parameter selection. We emphasize that our FIS is not meant to replace or to substitute for an experienced microbiologist; or laboratory investigations; on the contrary, we envisage that the FIS should be viewed as a decision aid for the busy emergency-department, microbiologist, particularly in times of epidemics or water diseases.

The Fuzzy logic inference system shows that the output-number of bacteria in water corresponding to any variation in the input parameters. The result of the Fuzzy program so far, is a numeric and symbolic terms of number of bacteria in water, using the input data from sensors and output data in the universe of discourse (safety, contaminated; more contaminated). if the number of bacteria in output is around 5 ($\log N = 5$), the water is contaminated; if it is less than 4, the water is safety; if it is more than 6 it is more contaminated.

REFERENCES