Intelligent Computational Modeling and Prediction of Coliform Growth in Tropical Lakes based on Hybrid Self Organizing Maps (SOM) and Fuzzy Logic Approaches

Sharifah Mumtazah Syed Ahmad¹, Mohammed Balkit Turki²
College of Information Technology, Universiti Tenaga Nasional (UNITEN)
Kajang, Selangor, Malaysia
smumtazah@uniten.edu.my¹
muhmed1984@yahoo.com²

Sorayya Malek³
Institute of Biological Sciences (ISB), Universiti Malaya, Kuala Lumpur, Malaysia
sorayya@um.edu.my³

Abstract — This paper describes the feasibility of applying a combination of intelligent approaches of self organizing maps and fuzzy logic in modeling and predicting the growth of total bacteria coliform in tropical lakes. The application of self organizing maps (SOM) in this study is to derive the membership rules for ecological variables such as temperature, pH and biochemical oxygen demand (BOD) that influence the growth of bacteria coliform in the water of Putrajaya (tropical) Lakes and Wetlands. The membership rules derived from the SOM are used as the bases to tune the intelligent fuzzy prediction module. The overall system has been trained and tested for its accuracy using a reliable database that provides samples of relevant parameters from the lakes and wetlands over a period of five years. The system has demonstrated prediction of coliform growth with up to 90.5% of accuracy.

Keywords-Kohonen Self Organizing Maps, Fuzzy Logic

I. INTRODUCTION

Coliform bacteria are a collection of related microorganisms that live and reproduce in large numbers in human and animal intestines [1]. They can also be found in areas where human and animal fecal matter may be present such as in soil, surface water, and plants. Though coliform bacteria may be relatively harmless, the presence of a group of fecal coliform in water which is known as E. coli (Escherichia coli) may cause diseases such as cholera, typhoid fever, dysentery etc. Thus, it is highly desirable to model the water quality by studying the growth of coliform and E. Coli in water content. Though much research work is carried out in emphasizing the modeling the quality of water resources that is used directly for drinking purposes [1,2 and 4], there are also increasing research efforts in modeling the water quality of natural water resources such as lakes and rivers [5 and 6].

Fuzzy Logic has been applied to assess the water quality of Goksu Stream [7]. In addition, an inductive model has also been developed using other artificial intelligence techniques for fecal coliform prediction and classification in surface waters. The AI techniques used include artificial neural networks (ANNs) and a fixed functional set genetic algorithm (FFSGA) approach for function approximation. FFSGA has been reported to be an effective technique that competes well with other complex techniques such as ANNs [8].

The Self-Organising Map [9, 10] is a topologically-based unsupervised clustering algorithm that has been successfully applied to a variety of ecological data sets for visualisation of relationships and representation of state changes in ecological systems [11] and stream ecosystems [12]. The SOM works by effectively creating a two dimensional spatial representations that maps selected external parameters (also known as input features) to their corresponding output abstraction of the total coliform count in a water sample. This technique is chosen mainly due to its ability to analyze, classify and visualize data with a high degree of complexity as demonstrated in [13].

Furthermore, as ecological data accumulates and increases in complexity, there is a need for artificial neural networks (ANN), and in particular, the self-organising map (SOM), in analysing ecological datasets. Thus SOM has been applied in the context of this study to assist in the extraction of propositional rules that highlight the relationship between the relevant input and output parameters.

Fuzzy Logic meanwhile is chosen mainly due to its capability to make decisions in an environment of imprecision, uncertainty and incompleteness of information [14]. This criterion is particularly useful in this scenario due to two limitations of the propositional rules extracted earlier by the SOM module. First, the graphical two dimensional SOM diagrams only allow for derivation of a rough estimation of rules; in other words rules with some level of imprecision. Secondly, the SOM only allows for extraction of rules, but it does not provide any former information on the prioritization of rules. Thus, a straightforward rule based prediction module may not be the best solution, since it requires prior knowledge of the prioritization of the rules. Fuzzy Logic is superior in this sense whereby it allows all possible rules that meet the criteria of the input parameters to be ‘fired’ simultaneously and later to be combined to arrive at a final decision in an effective manner.
Thus, this paper aims:
(1) to demonstrate the feasibility of applying Self Organizing Maps (SOM) to derive membership rules for modeling the water quality, particularly the total coliform growth with respect to the selected environmental factors such as water pH measurement, Biochemical Oxygen Demand (BOD) and water temperature;
(2) to develop an intelligent fuzzy predictive system for total coliform in order to demonstrate an effective forecasting computational module in predicting the water quality.

The organization of this paper is as follows: Section II describes in detail the experimental data sets used in this study. Section III presents elaborations on the usage of Self Organizing Maps (SOM) to derive the fuzzy membership rules, whereas Section IV describes the resulting fuzzy memberships and rules used in the intelligent prediction module. The experimental results are presented in Section V, followed by the conclusion of the paper.

II. EXPERIMENTAL DATA SETS

The data used in this study are from selected sampling points of Putrajaya Lakes and Wetlands, Malaysia. It is located in the south of the densely inhabited Klang Valley. The wetland covers an area of 650 ha including the large Putrajaya Lake with its marshes and wetland cells and has adopted a multi-cell design strategy. The multi-cell layout comprises of six defined wetland arms and lakes which are: Upper North, Upper West, Upper East, Lower East, Upper Bisa, Central Wetland and Putrajaya Lake [15]. Putrajaya Lakes and Wetlands water bodies are classified as oligotrophic to mesotrophic, which indicates that the water is relatively clean.

Limnological parameter data used for the purpose of this study is compiled over an extensive period of five years. Sampling procedures, which includes preservation for water quality parameters, were carried out in accordance with WHO-GEMS [16] and APHA [17]. The analytical methods for the measured parameters were adopted from a manual published by the American Public Health Association [17].

The data for this study is categorized into two different sets:

(1) **Data set A**: used for system training and deriving the membership rules. The data are used to build the water quality model by assisting in the extraction of the relevant propositional rules from the SOM module. The relevant rules are then used to develop the fuzzy prediction module.

(2) **Data set B**: used for system testing purposes for the fuzzy predictive system. This is to avoid obtaining biased results where a separate data set is fed into the fuzzy logic prediction module to effectively test its feasibility and accuracy.

III. SELF ORGANIZING MAPS (SOM) TO DERIVE FUZZY MEMBERSHIP RULES

The SOM is trained using training dataset A in order to derive fuzzy membership rules. Variables used are depicted in Table 1. The input variables are water pH, temperature and BOD which are mapped against the output variable total coliform count.

Water temperature is chosen as one of the input variables since it is well known to influence the microbial growth in water either directly or indirectly. Significant microbial activity in water is only found when the temperature of 15 degrees Celsius or higher. Whereas pH is a measure of the acidic or basic (alkaline) nature of a solution and a pH range of 6.0 to 9.0 is the suitable range for microbes to thrive. Biochemical Oxygen Demand (BOD) is important because it shows the amount of organic matter that is in the water. In other words, this is the measure of the amount of oxygen that would be consumed if microorganisms oxidized all of the organic matter in one litre of water.

**TABLE 1. CATEGORIZATION OF PARAMETERS INTO INPUT AND OUTPUT VARIABLE OF THE SOM**

<table>
<thead>
<tr>
<th>Division</th>
<th>Variable Category</th>
<th>Category</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>Temperature Physical</td>
<td>'C</td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>Chemical</td>
<td>'C</td>
<td></td>
</tr>
<tr>
<td>Biochemical Oxygen Demand (BOD)</td>
<td>Chemical</td>
<td>mg L⁻¹</td>
<td></td>
</tr>
<tr>
<td>Output</td>
<td>Total Coliform Biological</td>
<td>Cfu/100mL</td>
<td></td>
</tr>
</tbody>
</table>

First, the output variable total coliform species count is labeled into ‘Low’, ‘Medium’ and ‘High’ categories based on statistical distribution of the dataset. This step needs to be carried out before the training phase of the SOM which assists in interpreting the SOM diagrams for deriving the fuzzy membership rules. The thresholds that differentiate these clusters are as represented in Table 2.

**TABLE 2. CATEGORIZATION OF WATER QUALITY BASED ON TOTAL COLIFORM COUNT**

<table>
<thead>
<tr>
<th>Label</th>
<th>Threshold (Total Coliform Count) Cfu/ 100mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>&lt; 10</td>
</tr>
<tr>
<td>Medium</td>
<td>10 – 60</td>
</tr>
<tr>
<td>High</td>
<td>&gt;60</td>
</tr>
</tbody>
</table>

Figure 1 below illustrates the U-Matrix diagram generated by the SOM module based on the clustering category presented in Table 2. Visual inspection of Figure 1 reveals that the majority of the dataset falls under the high category for total coliform.
The rules to model and to predict the algal growth are then extracted by mapping the clusters described in Figure 1 into the input variables self organizing maps of temperature, pH and BOD as illustrated in Figure 2, 3 and 4 respectively.

**IV: WATER QUALITY PREDICTION THROUGH THE USE OF FUZZY LOGIC**

The rules extracted from the SOM are then used to tune the fuzzy prediction module. Since there are three input variables used in the study, the resulting Fuzzy Association Matrix can be illustrated in three dimensions of size 3 x 3 x 3, corresponding to the fuzzy memberships of ‘Low’, ‘Medium’ and ‘High’ for each input variable respectively. In this study, the categorization of the input variables into different categories as well as the determination of the fuzzy membership shape (either into half triangle, trapezoid or uniform triangle) are carried out based on the rough estimation of the SOM outputs. The membership functions for the input variables are as demonstrated in Figure 5, 6 and 7 respectively.
The resulting fuzzy rules can be summed up in Table 3, 4 and 5 below.

**TABLE 3. FUZZY RULES WHEN BOD IS ‘LOW’**

<table>
<thead>
<tr>
<th>pH</th>
<th>Temp</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4. FUZZY RULES WHEN BOD IS ‘MEDIUM’**

<table>
<thead>
<tr>
<th>pH</th>
<th>Temp</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 5. FUZZY RULES WHEN BOD IS ‘HIGH’**

<table>
<thead>
<tr>
<th>pH</th>
<th>Temp</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

V. EXPERIMENTAL RESULTS

A different data set, namely dataset B which was not used for SOM training was used to test the effectiveness of the Fuzzy Prediction Module. First, the total coliform count of the data in the testing of set B was labeled into the corresponding category of ‘Low’, ‘Medium’ and ‘High’ based on the defined thresholds as described previously in Table 2. Then the normalized input variables of water pH, temperature and BOD measurements were fed into the fuzzy prediction module. The output prediction state of the fuzzy system was recorded and compared against the actual coliform state whereby a miss-match of both data would give rise to an error count. The accuracy of the system was determined by taking the percentage of the total number of accurate predictions over the overall tested data. In this study, the fuzzy prediction module yielded a performance accuracy of 90.5%.

VI. CONCLUSION

The objective method of computing total coliform in order to present a concise picture of overall water quality trends is useful as a tool in management of water quality for lakes and wetlands. The approach outlined in this paper using the hybrid approaches of SOM and Fuzzy Logic to perform water quality modeling and Coliform prediction of tropical lakes and wetlands of Putrajaya could be a better representation of a dynamic system, and thereby providing a new dimension of lakes and wetlands management in Malaysia. The authors believe that the approach proposed in this study, if used intelligently, could be an effective technique for some of the environmental management issues.

ACKNOWLEDGMENTS

We would like to express our acknowledgement to Environmental, Lake and Wetland Division, of Perbadanan Putrajaya for providing us with the relevant data for this study.

REFERENCES


[3] Xiao Sheng Shen; Hui Juan Yu; Yun Hua Hui; Run Run Gu; Wei Min Quan; ChengChu Liu, “Seasonal Variations of Total Coliform and Fecal Coliform in the Lagoon Seawater of Shanghai Jinshan City Beach in North Hangzhou Bay”, Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering, 2008


