BUILDING UP AIR POLLUTION PREDICTION MODEL IN HO CHI MINH CITY USING GIS AND NEUTRAL NETWORK TECHNOLOGY

Bui Ta Long[1], Luu Minh Tung[1], Le Thi Quynh Ha[1], Duong Ngoc Hieu[2].

[1]Institute of Environment and Resources, Vietnam National University of Hochiminh city
142 To Hien Thanh, District 10, HoChiminh city, Vietnam
Email:buita@hcmc.netnam.vn

[2]University of Technology, Vietnam National University of Hochiminh city
268 Ly Thuong Kiet St., Dist. 10, Hochiminh city.

ABSTRACT

Presently, one of the missions on top priority of authorities at all level in Ho Chi Minh City is the analysis and prediction of air pollution situation in the most populated city of Vietnam to date. To address this problem, it is essential to apply the latest results from variety of technological and scientific investigation works, amongst is information technology, and especially is artificial intelligence approaches in associated with Geographical Information System (GIS). Investigation result originated from the applications of artificial intelligence approaches will enable the circumstance for the appearance of numerous applications on the way toward the use objectives those whom are experts on mission at Resources and Environment Services.

In this work presents the results derived from execution of scientific and technological research project on Ho Chi Minh City level carried out in 2 years 2004 – 2005. The key result of this project is the software product entitled SAGOCAP (Contaminants in the Air Prediction model for SAi GOn) on GIS application and neutral network technique helped to manage monitoring data on air quality at 9 automatic monitoring stations located in Ho Chi Minh City area as well as it is to support air quality prediction at these stations using neural network technique. Through this report, the authors hopes to share the idea and to assert the artificial intelligence integration method with GIS as well and it is particularly effective on the sector of environment decision making.

1. INTRODUCTION

In the context of impact caused by human – being socio-economic activities on ceaselessly increasing that the most important issue is to organize the environmental situation monitoring and continuously monitoring the change of environment and to determine the their tendency of change. In the face of environmental imperativeness in Ho Chi Minh City, since 2000, Ho Chi Minh City Service of Technology and Environment has installed the automatic air pollution monitoring system of which has ever satisfied international standard in Vietnam.

At the time of installing the continuous monitoring system to date, such monitoring data has uninterruptedly been increasing. Though, so many approaches investigated and applied to analyze this information flow, it might assert that there is virtually no any tool used for information analysis on effective (as of temporal parameters), rapid and particularly capable of bring out predictions so far. It is needed to underline that just not in Vietnam where has to confront to such issue. Building up a technology that enables a sole decision making on the basis of huge information flow in a very limited time scale is the great concern of a lot of developing countries worldwide. It was because many countries in the world have taken much time to build up tools on the basis of advanced and powerful technologies such as artificial intelligence. Hence, the necessity on building up such a tool is that:

- To get to know the relationship between various parameters without the expressly interdependence between them is a very difficult work, even in presence of most proficient experts. In addition, it is needed to note that the sometimes, subjective thought of expert leaded to the prejudiced prediction. In the situation of time limited, experts may commit an error.

- The uninterrupted increase of information current and so difficult to expert (even in the assistance of modern database management system as well as other computer methods that help to seek information)

- In the environmental prediction problem, the dependence of pollution with other parameters is obvious but to imitate this dependence could not be achieved by old tools. Another importance is that could be said that this information space is currently too wide and beyond the control, information is so diversified, thus, it is hard to apply conventional information processing methods.

- Neural network was proved fairly strong and effective in reality in prediction problems and data analysis. Neural network applied to prediction pollution is a real demand helping environmental management agencies to figure out appropriate policies and decisions soon.

Thence, the theme entitled "*Research for building up air pollution prediction module in Ho Chi Minh City through neural network approach*" is carried out in line with the goal of forming a closed tool as from information classification, automatic information processing and winding up by information analyzing procedure and pollution prediction. The aim of this theme is the application of neural network technique to predict air pollution at automatic monitoring stations in Ho Chi Minh City based on the temporal monitoring figure range.

2. BUILDING UP NEURAL NETWORK FOR AIR POLLUTION PREDICTION

According to Ho Chi Minh City service of Resources and Environment, currently there are 9 automatic air pollution continuous monitoring stations in Ho Chi Minh City located at: Service of Science and Technology, Hong Bang School, Binh Chanh District Education Bureau, Thong Nhat Hospital, Quang Trung Software Park (District 12), Zoological and Botanic garden (District 1), District 2 People's Committee, Urban administration bureau (Thu Duc District), Tan Son Hoa (Phu Nhuan District). Contaminations are measured at 24/24 including PM_{10}, SO_2, NO_X, CO, O_3**.** All Survey equipments are modern at high accuracy.

2.1 Designing neural network system used for air pollution prediction.

There are 3 steps on building up neural network model: data pre-processing, network training and prediction (result output) and respectively to foregoing 3 steps, neural network model for air pollution includes 3 main modules as follows:

Data preprocessing module

- Data of our survey from monitoring stations are recorded in the form of Excel file then data preprocessing module will convert such data into database SQL Server 2000.

- Apart from data we got from monitoring stations, usually there are a lot of errors that such errors caused by the incapability of survey equipment. Such values therefore are eliminated. This is the step of interference elimination.

- After completing interference elimination, we move to data highlight stage. Data highlight means that we have to highlight the property of pollution data prior to entering them into neural network, namely in air pollution prediction problems at automatic stations, we may realize that this data is cycle nature as of hour of day, day of year and month of year. So, hour, day, month values are selected here as the role of 3 signals to enter into neural network.

Network training module

This module enables user building up network architecture and configuration of parameters. Then, such module employs processed data file to train network. This step is

divided into 2 smaller ones are training and verification. The network training step employs 90% of available data. Verification step employs approx 10% of the rest to verify the quality of the network to get to know if the network is on good running. After that, we evaluate error of neural network that we have completely trained.

Prediction module

This module enables user predict needed results.

2.2 Applying backpropagation technique into air pollution prediction problem

In order to build up air pollution prediction model as of continuous monitoring data range under neural network technique, in backpropagation algorithm application theme.

The steps of construction include structure determination as well as the selection of binding important number training algorithm. After training, we take the role model of prediction function. Based on this prediction function, we may determine value of contaminations.

Structure

Today, multilayer feedforward network architecture is mainly used on prediction application.

- Input layer: number of neuron of input layer is equal to direction of vector **X**. It is the meteorological condition number + outstanding data number + temporal factor.

- Hidden layer: number of neuron may change at own discretion

- Output layer: Number of neuron is the value of contamination concentration that we need to prediction.

Sample data selected in the theme

Sample data set of training process in this theme is set $X = \{(\mathbf{x_i}, \mathbf{d_j(x_i)})\}^{N}_{i=1}$

Where:

- $\mathbf{x_i}$ – is the vector of meteorological condition values (Such as Max Temperature, Min Temperature, wind velocity, wind direction, head, radiation intensity), time, value of contamination concentration after being highlighted.

- $\mathbf{y}_j$ (j =1, 2,…J), value of contamination concentration needed to be predictioned.

- $\mathbf{d_j(x_i)}$ – is the prediction value

There are a lot of approaches to highlight data, in this theme used n past data approach to prediction data n + 1. Table 1 shows rules for highlighting in this theme (n=3).

Table 1.Way of selecting value highlighted for predictioning.

Day	Hour	Concentration value of contamination selected	
1/1/2003	1	12.4	
1/2/2003	1	11.5	**Input**
1/3/2003	1	4.6	
1/4/2003	1	6.5	**Output**

Carrying out back-propagation algorithm in the theme

As presented in previous chapter, training process on neural network using back-propagation algorithm includes 3 stages: feed-forward stage of input signal, in this stage, forward procedure is called out to execute forward-propagation stage – constitutes the value

determining of the mapping on the basis of existing coefficient), different error computation stage between desired value and value really gained from neural network and the ultimate stage is the adjustment of connection weights

- Feedforward stage: Each neuron in input layer gets the input signal and sends this signal to all neuron in the hidden layer. Each neuron in the hidden layer gets signal from the input layer will calculate the output via activation function and sends the calculation result to all neuron in output layer. Neurons in the output layer will carry out the same calculation with neurons in the hidden layer.

- Backpropagation stage: During the training process, after completing the calculation of value at the output, each neuron in output layer will compare the gained result to desired value to determine the different error between them called error signal. Information about error signal will be used to adjust connection weights on the direction as from the output layer to the hidden layer and continue doing so until the last one in input layer.

- Adjustment stage: It is the final step of training process. There are two ways for adjusting connection weights: Batch form way refers to the update process of connection weights occurred as all training samples processed. Incremental way refers to the update of weight occurred soon after a sample was completely processed. However, incremental way is used much more in prediction problem.

2.3 SAGOCAP tool – automatically computing air pollution prediction under neural network technique model

Within the scope of this theme, the authors built up SAGOCAP (**C**ontaminants in the **A**ir **P**rediction model for **SA**i **GO**n) software applying of GIS and neural network technique to carry out air pollution prediction function.

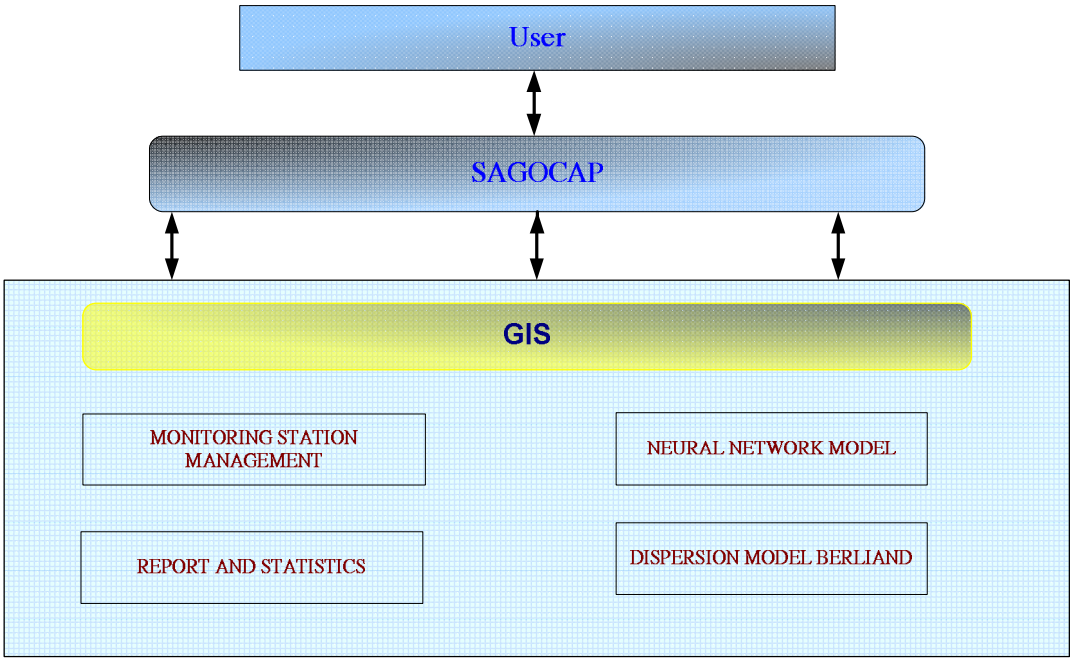

Figure 1. General schematic structure of SAGOCAP

SAGOCAP version 1.0 enables managing subjects related to quality of air environment: monitoring figures on air environment pollution from 9 automatic gauging stations, surveyed meteorology figures and enables predictioning air pollution. These subjects are fixed to numerical map on the Hochiminh city ground. The overall sketch of SAGOCAP shown in Figure 1, SAGOCAP has following key functions:

- Central administration and unifying the monitoring data from 9 automatic monitoring stations, the quantity of such stations may increase, thus SAGOCAP enables users opening new station as in the need;

- Conducting statistical functions, report on the basis of figures entered into database. The batch sketch of this function is shown above

- Conducting prediction functions as per neural network method. The working outline of this module is shown on fig. 2.

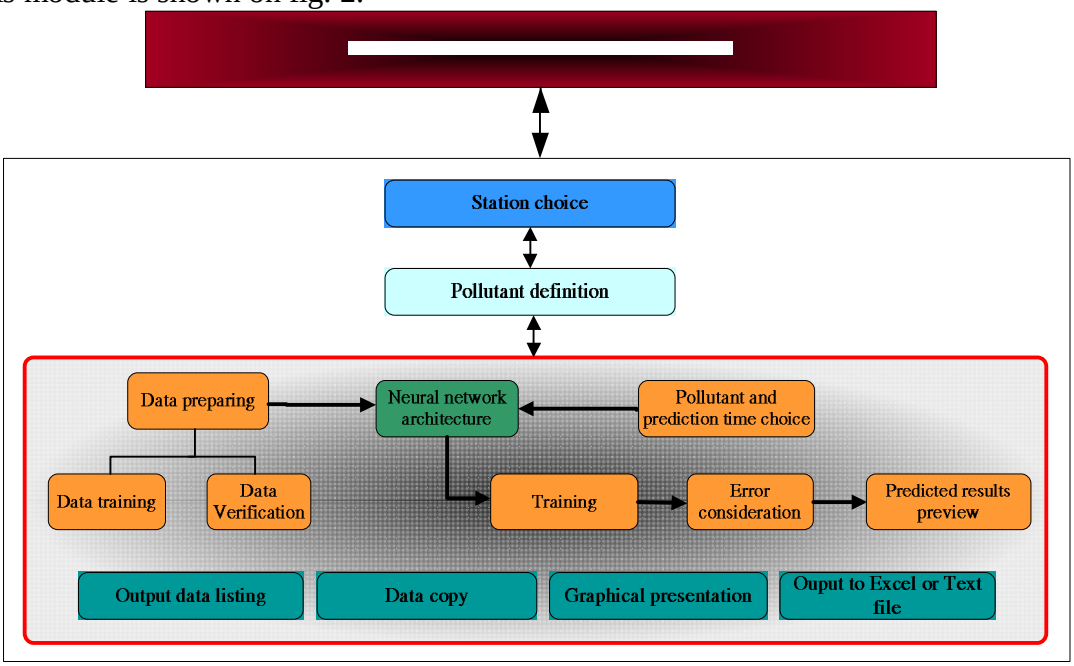

Figure 2. Working schedule of prediction model in SAGOCAP

Figures used in this theme provided by environment protection branch of Ho Chi Minh City as per contract. From Excel format file is converted into format SQL. then, the user needs to select the date needed to be predictioned as well as station and substance needed to be predictioned (there are 9 automatic monitoring stations in Ho Chi Minh City).

The next will be the self training process over available figures. This study process is kept continuing until an acceptable error attained and then stop. It is aim to test capability of SAGOCAP1 version 1.0, authors employed the dataset of 2003 for self training process of SAGOCAP1 and then outputting prediction results of 4 days in early 2004.

2.4 Result of air pollution prediction using neural network model

Within the performance framework of this theme, from June, 2005 to December, 2005, SAGOCAP product was installed at the bureau of monitoring and evaluating environment pollution subordinated to Ho Chi Minh City environment protection branch. The neural network model testing was just conducted by staff of environment pollution monitoring and evaluation bureau upon the contract on fixed rate professional hiring between the theme head and Ho Chi Minh City environment protection branch (certified by Institute of application mechanics, the governing body of the theme), below presenting certain results were initially received.

SAGOCAP is used for prediction goal at monitoring locations: Thong Nhat Hospital (PM_{10}, CO), HCMC Service of Science and Technology (0_3), Binh Chanh district education department (presently belonging to Binh Tan District) (NO_x), HCMC Zoo (NO_x), District 2 People's committee (NO_x, PM_{10}). Comparison result between value predictioned by SAGOCAP and real gauged value presented in form of chart below at some stations /fig. 3/.

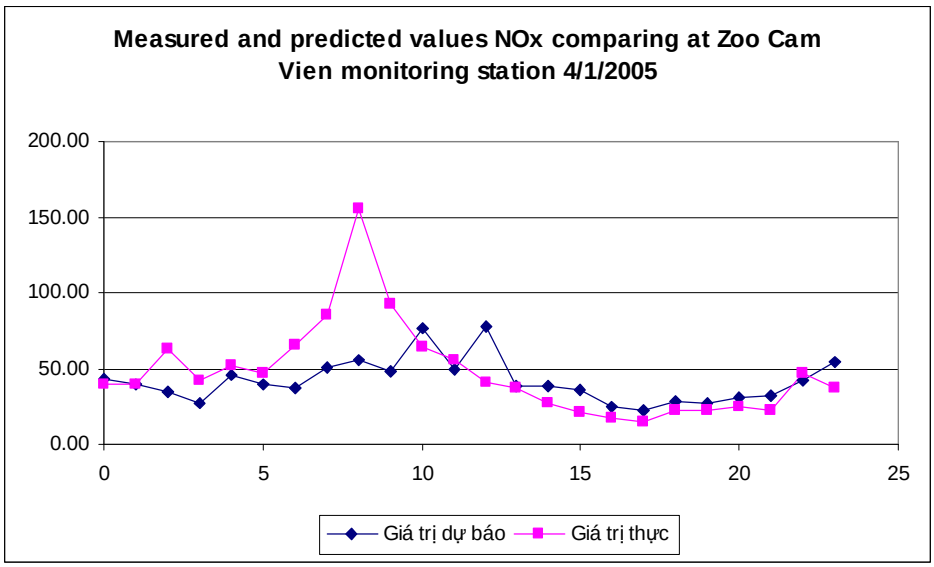

Figure 3. Measured and predicted values NOx at Cam Vien zoo monitoring station at 4/1/2005

4. CONCLUSION AND PROPOSAL

The main result of this article is a tool for promptly and analyzing temporal changing monitoring data and particularly it may expose air pollution prediction. The result of this research was applied to prediction air pollution in Ho Chi Minh City. Informatics tools built in this work enables resolving tasks of environmental figure administration in current stage.

The shortcoming of this research is just the issue of figures. Presently, temporal figure range provided by automatic monitoring gauge being in serious error that interrupted such figure range. Through analysis, the author found out that most of data were interrupted and yet the degree may vary on the basis of every station or every substance. It is difficult to improve the accuracy of the prediction.

Artificial intelligence technique in general and neural network technique in particular plays a great role on handling socio-economic issues and environment as well. The result of this theme enables affirming a possible prospective of on solving abovementioned problem through advanced technology.

5. ACKNOWLEDGEMENTS

The present study is financially supported by grant from National program by natural sciences code 7.197.06. The authors are thankful to Department of Sciences and Technology, Hochiminh city for providing a number of good quality databases to carry out the study.

6. REFERENCES

[1]. Long Bui Ta at el., 2001. The GIS based simulation computer system for analysis, visualization, and prediction traffic air pollution. Proceedings of the IASTED International Conference Applied simulation and modelling. Sep. 4 – 7, 2001, Marbella, Spain, pp. 359 – 364.

[2]. Ashutosh Tewari, Genetic Adaptive Neuron Network to Predict Biochemical Failure After Radical Prostatectomy: A Multi-institutional Study, Mary Ann Liebert Inc, 2001

[3]. T.T. Chow, Z. Lin and C.L. Song, Applying Neuron Network and Genetic Algorithm in Chiller System Optimization, 2001