ANALYSIS OF FLOW METERS READING IN FUKUOKA CITY WATER SUPPLY NETWORK USING SELF-ORGANIZING MAP

HAYTHAM AWAD, AKIRA KAWAMURA AND KENJI JINNO

Institute of Environmental Systems, Kyushu University

6-10-1 Hakozaki, Higashi-ku, 812-8581, Fukuoka, Japan

The water distribution regulation system of Fukuoka City is a system in which motor valves are operated by remote control while pressure gauges and flow meters attached to distribution pipes are monitored. One of the main objectives towards the wise management of this system is to regulate pressure in all the network nodes between lower and upper target values. With the ever-increasing complexity of the city-wide distribution pipe network, motor valve operations to regulate pressure and flow came to depend more and more on the experience and skills of operators. For this reason, an improvement of valve operations support should be done based on valve operation planning for flow and pressure regulation, and the operation knowledge database which is constructed on the basis of past experience in order to prevent the events of pressure regulation falls outside the required target pressure range, feeds the different network nodes with the required water demand and also to reduce the effort of operators. Therefore in this study an analysis of the existing flow meters reading of a certain block within the city district is performed using an unsupervised class of Artificial Neural Networks (ANN) named Self-Organizing Maps (SOM). Two different models are constructed; the first is used for long term operation with the assistance of three years of hourly telemetry data while the second model is performed using two days of one minute telemetry data for short term operation. Both models treat the valve operations rules and the resulted water pressure by an indirect way. Results show that SOM is an efficient tool in clustering the different complicated operational patterns of valves, visualizing the huge amount of telemetry data, detecting the patterns which need future improvement and could present good alternative solutions for improving future valve operational support.

INTRODUCTION

Fukuoka City is poor in water resources with no large rivers within the city district. For this reason it has taken various measures towards the creation of a “water-saving city”, a city in which valuable water will not be wasted. One such measure has been the installation of the water distribution regulation system. With this system, round-the-clock centralized monitoring of data from pressure gauges and flow meters attached to distribution pipes is carried out, while on the basis of this data, pressure and flow within the entire city district is regulated by the remote operation of motor valves. By putting this system into operation, it has been possible not only to supply water evenly to consumers, but because water pressure can be regulated appropriately it has also been possible to reduce excess pressure and conserve water [1].

One of the main objectives towards the wise management of this system is to improve the existing motor valve operations which aim to find the optimal electric openings of different motor valves to achieve two targets; feed consumers with their requirements of water demand and try to regulate water pressure in the entire water supply network between upper and lower values. The lower pressure target value is assigned because consumers want to receive water with adequate pressure while the upper target value is introduced for the purpose that higher pressures cause an increase of the amount of leaked water from the network.

Available models that explore pressure regulation problem through optimal control valve settings could be divided into two categories. In the first type, a mathematical statement is used as an objective function to be minimized. This mathematical statement could be presented in the form of network pressure regulation [2] or as the total amount of leaked water from the network [3]. The main drawback of using this category of models is its computational time. The objective function of all available models of that class deal directly with the pressure regulation problem from an optimization point of view which required a computational time depending on the water network size, the number of variables to be optimized and the used optimization method and in most cases required a network simplification method which is considered as another optimization problem. The second category of the pressure regulation models are based on extracting several useful operational rules based on the previous knowledge and experience, analyzing past recorded data and the present experience and skills of operators. The main advantage of this kind of models that it could be used as on-line operational model providing us with the most recommended values for electric motor valves set and providing us with the system response for those recommended values.

In order to extract several useful operation rules, this study present a classification analysis of past recorded flow meters reading cases of Block 12 of Fukuoka City water supply network using an unsupervised class of Artificial Neural Networks (ANN) named Self-Organizing Maps (SOM). The analysis are performed using two sets of flow meters telemetry data; three years with one hour interval and two days with one minute interval.

ANNs are already quite commonly used in several applications of water resources engineering as an alternative for conventional computational models. ANN used an ensemble of input-output patterns to model a map from an input layer to an output layer, which is considered in this case a supervised type of ANN. SOM are considered also a class of ANN, but trained in an unsupervised way. That means SOM do not require an associated output (target) for each input pattern during training. The process of SOM classify the input patterns to different groups based on measuring similarity between input patterns, with no or little knowledge about the structure of used data.

The purpose of using SOM in this study is to test its ability of clustering the different cases of flow meters reading, visualizing the huge amount of telemetry flow meters data and detecting the groups which need future improvement. Finally, using the results obtained from the SOM, we could suggest good alternative solutions for improving future valve operations.

SELF-ORGANIZING MAPS (SOM)

The SOM is relatively a simple unsupervised neural network used for the categorization of input patterns into a finite number of classes. SOM consists of two layers units, the input units are a one-dimensional array which provides simulation to a usually two-dimensional array of output units and all units in the input layer are fully connected with the units in the output layer. The neurons of the output layer which is preferable to be arranged in two dimensional grids for better visualization are connected to adjacent neurons by a neighborhood relation dictating the structure of the map. The arrangement of the output layer neurons are usually distributed in rectangular or hexagonal arrangement. Generally it is preferable to use the hexagonal lattice, because it does not favor horizontal and vertical directions as much as rectangular array [4].

When an input vector x is sent through the network, each neuron k of the output layer computes the distance between the weight vector w and the input vector x. Among all the output neurons, the so-called winning unit or Best-Matching Unit (BMU) is determined by the similarity between the weight vector w on that unit and the input vector x. For an input vector x, the BMU is determined by

$$\|x - w_c\| = \min_i \left\{ \|x - w_i\| \right\} \tag{1}$$

where the subscript c refers to the winning unit (BMU), $\|\ldots\|$ is the distance measure and i refers to all units in the competition layer. Accordingly, a second winning unit will be determined with respect to the second input vector, and so forth. At the end of competition only one unit in the competitive layer wins in corresponding to one input vector.

The training is usually performed in two phases. In the first phase, relatively large initial learning rate and neighborhood radius are used. In the second phase both learning rate and neighborhood radius are small right from the beginning. This procedure corresponds to first tuning the SOM approximately to the same space as the input data and then fine-tuning the map. After some training steps, the SOM will arrange high-dimensional input data along its two-dimensional output space such that similar inputs are mapped onto neighboring regions of the map which means that the similarity of the input data is preserved within the representation space of the SOM. To ensure that all variables of any input vector x receive equal attention during the training process, it is important to normalize the input vector to unit length before the training steps.

To measure the ability of SOM to arrange the different input vectors through its two-dimension grid, usually two evaluation criteria could be applied to measure the quality of SOM; resolution and topology preservation. For identifying and measuring the resolution of the SOM, we compute the quantization error [4] which is the average distance between each data vector and its winning unit (BMU). The topographic error which used to present the accuracy of the training map in the preserving topology is also calculated.

This error represents the proportion of all input data vectors for which first and second BMUs are not adjacent for the measurement of topology preservation. The topographic error can be calculated as follows [5]:

$$\varepsilon_t = \frac{1}{N}\sum_{k=1}^{N} u(x_k) \tag{2}$$

where N is the number of input vectors; $u(x_k)$ is 0.0 if the first and second BMU's of x_k are next to each other, other wise $u(x_k)$ is 1.0.

In this paper we used SOM to classify different cases of flow meters reading, each represented by one vector. After classification we used two methods to cluster the obtained SOM to several main groups. First, we applied the method of unified distance matrix [4] (*U*-matrix) which is based on calculating the distance between adjacent SOM units. After calculated the *U*-matrix it could be visualized on a special color map size and then we could detect the different clusters using a color scale display on the map.

Another method that we have applied to select the best number of groups is computing the Davies-Bouldin Index (DBI) [6]. The smallest value of DBI represents the best number of groups which indicate the best clustering. Therefore, a choice is made concerning the number of clusters at which the DBI attains its minimum value. With the results obtained from clustering we could detect the well operated groups and the groups that need future improvement.

CASE STUDY AND DATA USED

Block 12 of the city network is selected as a case study. In this block, there are 20 motor valves, 7 flow meters, and 11 pressure gauges. All flow meters attached to the pipes of this block are connected to the main inlets and outlets and a valve is connected adjacent to each flow meter in order to control the flow entering or leaving the block.

The analyzed data in this study are based on two data sets. The first set represents hourly data for all telemeters since 1st April 1998 to 31st March 2001. This makes the total number of data for each telemeter 26304. The analysis of flow meters reading is based on 24062 vectors out of the 26304 vectors because there are 2242 missed vectors of flow meters data set. The second data set contains one minute data for a randomly selected two days (Saturday and Sunday, 9th & 10th of Nov. 2002).

SOM ANALYSIS

Each flow meter of Block 12 has its own time series operation during which the valve opening varies between a minimum and maximum values. To get an idea about the different flow meters readings characteristics, Figures 1A and 1B shows a box-whisker plot for the three years and two days data sets, respectively.

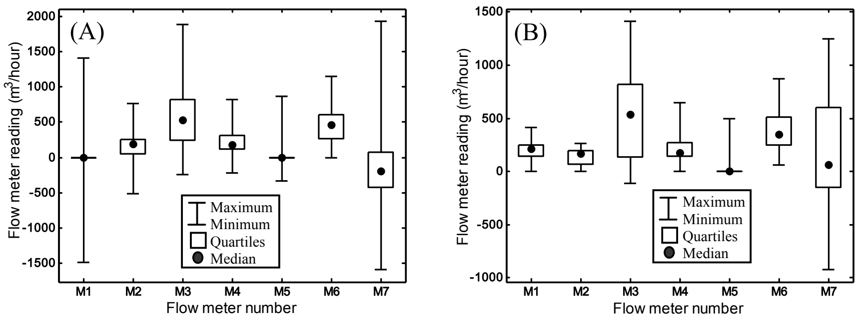

Figure 1. Box-whisker plots for the 7 flow meters, (A) 1st Data set and (B) 2nd Data set.

Considering that the size of SOM will affect directly the classification of input vectors, therefore a suitable size of SOM should be used. Different map size has been evaluated by calculating both topographic and quantization errors. Using the three years data set, Figure 2 shows contour map for both kinds of errors for all map sizes varies from 2 to 30 neurons. It is clear from both error contour maps that the size of the map has a strong effect on the distribution efficiency of input data vectors to the different neurons of the map. In general, increasing the map size will increase the topographic error while brings more resolution into mapping when the quantization error decreases. In Figure 2, the topographic error increase rapidly near both axis and the quantization error has concave shape near the diagonal line which starts from the origin. The map size selected to present the different classification of flow meters is 21×23 . At that size the topographic and quantization errors equal 3.87 and 0.9045, respectively. That's mean that there is only 931 vector in which the first and second BMU aren't adjacent.

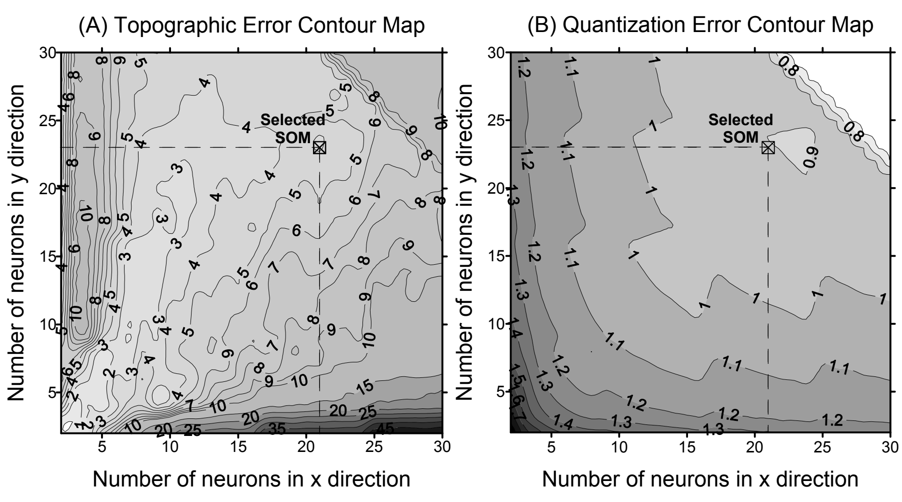

Figure 2. Contour maps of topographic and quantization errors for the 3 years Data set.

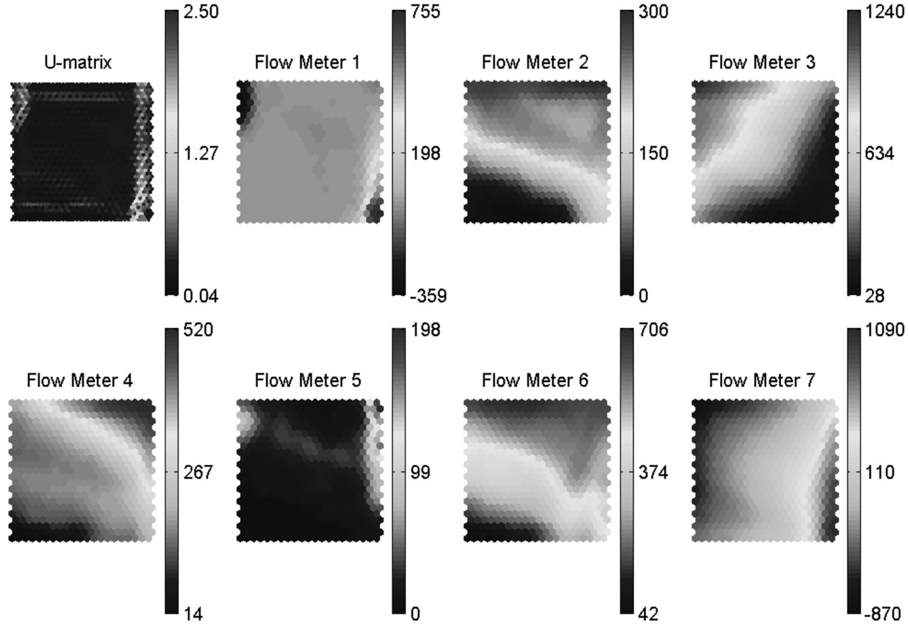

Figure 3. Visualization of (*U*-Matrix) and flow meters reading in the trained SOM

After selecting a map size of 21×23 the map is trained with the three years data set. Figure 3 shows a representation of all data components on the trained SOM. One of the good advantages upon classifying different input vectors by SOM that if any vector contains missing data the SOM could deal with this vector during the training process; each corresponding weight for the missing data vector will be neglected temporally from the calculation. Focusing on any unit of the map (for example the upper left one), we will see that M1 represents a reading of -350 m^3/hr up to M7 which correspond to -870 m^3/hr. The same procedure could be carried on other units. To compare the presentation of SOM components with the box-whisker plot of flow meters shown in Figure 1A. Both presentations give a good indication about different flow meters characteristics. When presenting an operation rule it is difficult for the box-whisker plot to efficiently represent the flow meters operations like the presentation of the SOM shown in Figure 3. Figure 3 shows also the *U*-matrix of size 41×45 where scale color bar on the right represents the differences between SOM units; each unit in the *U*-matrix map represents weight difference between two adjacent neurons of the competitive SOM layer. Highest value represents big difference and then the boundary between groups; if several adjacent units have same colors then they are located in the same group. To accurately estimate the main groups of flow meters readings, we compute the DBI for clusters number varies from 2 to 20. The minimum DBI is 0.702 recorded at 14 clusters. Thus, flow meters data were classified into 14 groups where each group represents a union of SOM neurons.

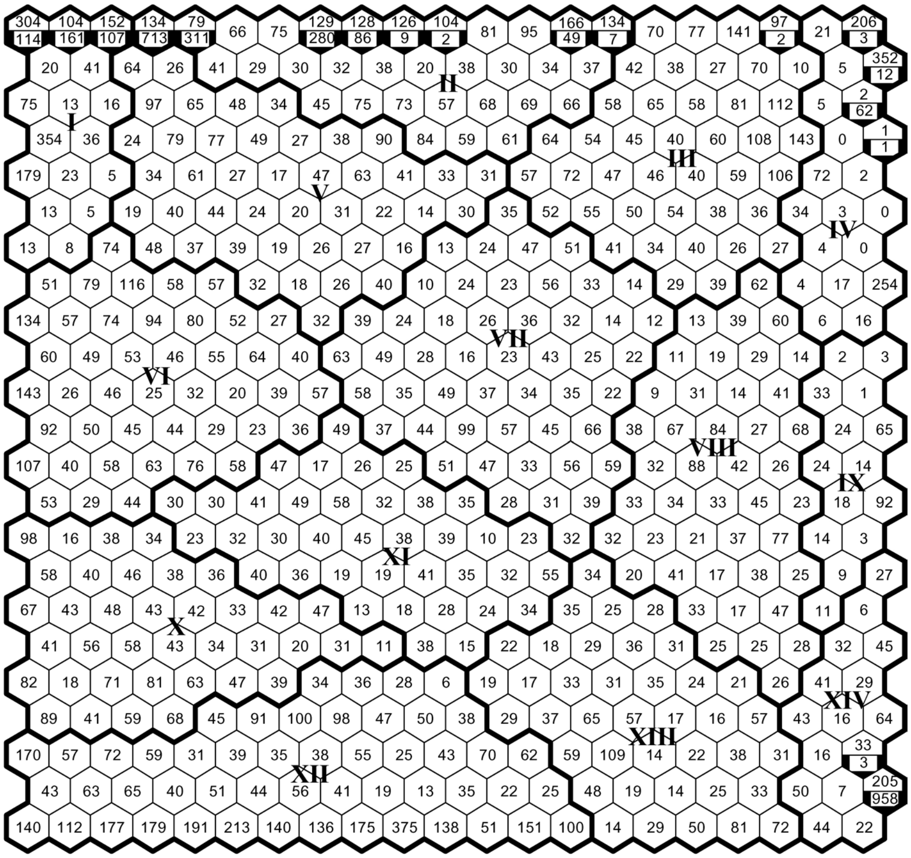

Figure 4. Clusters of the trained SOM units using the DBI and *U*-matrix methods.

Figure 4 shows the clusters of the trained SOM units. The number written inside each unit indicates the total number of hits associated with those units or the location of BMU each vector of input data. Boundaries of different groups mentioned in Latin numbers are set using the DBI and *U*-matrix methods. Distribution of the 24062 hit varies between a minimum value of 313 hit recorded at group IX to 4124 hit recorded at group XII. SOM could be considered as good tool for the presentation of highly dimensional data. It presents the characteristics of each data component and also presents the relation between different components. For the two-day one-minute data set a similar map could be obtained but in this study we present the location of the different 2880 hit of this data set in the resulted SOM of the three years data set. The second data set occurs only in five groups as shown in Figure 4 where the number of hits is recorded in lower area of each unit. The distribution of hits varies between 2 hits recorded at group III and 1457 hit recorded at group II.

SOM could be used for improving the existing valve operation support functions by using the trained map shown in Figure 4. Each unit in this map represents an embedded representation of the real water demand of the network. In this map any unit in the SOM presents both cases in which the pressure is well regulated and also unregulated. The corresponding valves opening set is also recorded for all those cases. After training the SOM, a simulation step is used to separate the unregulated pressure cases from the regulated ones according to their flow meters reading. Based on those classifications the appropriate electrical motor valves setting for the well pressure regulation events are used for the unregulated ones. This methodology offers good alternative solutions for improving future valve operational support functions.

CONCLUSIONS

This paper presents an analysis of flow meters reading in Block 12 of the Fukuoka City water distribution networks. The analysis has been performed on two telemetry data sets by using an unsupervised class of artificial neural networks named Self-Organizing Maps (SOM). The SOM has been applied for understanding the characteristics of flow meter readings for a set of three years with one hour interval of telemetry data. It has shown a high performance in visualization and abstraction of flow meters reading comparing to traditional methods.

The trained SOM efficiently classified the different operational rules and displayed all data components characteristics. With the assistance of the unified distance matrix (*U*-matrix) and the Davies-Bouldin Index (DBI) the trained SOM has been clustered into 14 main groups. The characteristics of each group could be analyzed to determine its effect on pressure values recorded at observed nodes. A simulation for a short term operation of telemetry data set has been performed on the resulted map and an application for improving future motor valve operation has been suggested.

REFERENCES

[1] Fukuoka City Waterworks Bureau, *"Water distribution control center"*, Publication of the Fukuoka City Waterworks Bureau, (2001).

[2] Savić, D. A. and Walters, G. A., "An evolution program for optimal pressure regulation in water distribution networks", *Engineering Optimization*, Vol. 24, No. 3, (1995), pp 197-219.

[3] Jowitt, P.W. and Xu, C., "Optimal valve control in water distribution networks", *J. Water Resour. Plng. And Mgmt., ASCE*, Vol. 116, No. 4, (1990), pp 455-472.

[4] Kohonen, T., *"Self-Organizing Maps"*, 3rd edition, Springer-Verlag, Berlin, (2001).

[5] Kiviluoto, K., "Topology Preservation in Self-Organizing Maps", *Proc. of the International Conference on Neural Networks (ICNN'96),* (1996), pp 294-299.

[6] Davies, D.L. and Bouldin, D.W., "A cluster separation measure", *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 1, (1979), pp 224-227.