RESERVOIR WATER RELEASE DECISION MODELLING

Wan Hussain Wan Ishak[1], Ku Ruhana Ku-Mahamud[2], and Norita Md Norwawi[3]

[1]*Universiti Utara Malaysia (UUM), Malaysia, hussain@uum.edu.my*
[2]*Universiti Utara Malaysia, Malaysia, ruhana@uum.edu.my*
[3]*Universiti Sains Islam Malaysia, Malaysia, norita@usim.edu.my*

ABSTRACT. Reservoir water release decision during emergency situations typically, flood and drought is very crucial as early and accurate decision can reduce the negative impact of the events. In practice, decision regarding the water release is made by experience reservoir operator. During emergency such as heavy upstream rainfall that may causes massive inflow into the reservoir, early water release cannot be done without the attendance and knowledge of the operator. Additionally, the operator has to be very certain that the water released will be replaced with the incoming inflow as maintaining the water level at the normal range is very critical for multipurpose reservoir. Having this situation every year the reservoir operation record or the log book has become knowledge or experience rich "repository". Mining this "repository" will give an insight on how and when the decision was made to release the water from the reservoir during the emergency situations. The neural network (NN) model was developed to classify the data that in turn can be used to aid the reservoir water release decision. In this study NN model 8-23-2 has produced the acceptable performance during training (93.94%), validation (100%) and testing (100%).

Keywords: reservoir water release decision, neural network, repository mining

INTRODUCTION

Reservoir dam is one of the defence mechanism for both flood and drought disasters. The use of dam for flood mitigation aims to impound water in a reservoir during periods of high flow in order to maintain safe downstream discharges (Smith and Ward, 1998). The opening of the dam's spillway gate must be adequate to ensure that the reservoir capacity will not over its limits and the discharges will not cause overflow downstream. During drought, the reservoir needs to impound water and release adequately to fulfil its purposes.

In both flood and drought situations, decisions regarding the water releases are made in accordance with the available water, inflows, demands, time, and previous release (Jain and Singh, 2003; Hejazi et al., 2008). The decision includes determining the quantities of water to be stored and to be released or withdrawn from a reservoir under various conditions (Wurbs, 1993). However, reservoir operation during these two situations is critical as it involve different objectives and purposes, thus required different operation rule. Moreover, these situations are not static where it changes as the subsequent to the climate changes (Hejazi et al., 2008). The relationship between the water release and the hydrologic information is nonlinear (Labadie, 2004; Hejazi et al., 2008) and there is a strong tie between them (Hejazi et al., 2008).

In this paper temporal data mining specifically sliding window technique is proposed to extract temporal data from the reservoir operation record. The backpropagation (BP) neural network (NN) was then constructed to learn the temporal pattern and perform the classification. The performance of the NN is measured based on the classification accuracy and the square error.

In the next section, an overview of temporal data mining is given, followed by an overview of NNs. The methodology of this study is presented in the research design section. The findings are presented in the findings section followed by the discussion and conclusion of the study.

TEMPORAL DATA MINING

Data Mining (DM), an activity that extracts some new nontrivial information contained in large databases (Laxman and Sastry, 2006), is a part of knowledge discovery in databases (KDD). Dunham (2002) defined DM as the use of algorithms to extract the information and patterns derived by the KDD process. Dunham divide DM tasks into eight categories typically, classification, regression, time series analysis, prediction, clustering, summarization, association rules, and sequence discovery.

Temporal data mining is branch of DM research. According to Laxman and Sastry (2006) temporal data mining is concerned with data mining of large sequential data sets (data that is ordered with respect to some index). Lin et al. (2002) defined temporal data mining as *"...a single step in the process of Knowledge Discovery in Temporal Databases that enumerates structures (temporal patterns or models) over the temporal data, and any algorithm that enumerates temporal patterns from, or fits models to, temporal data is a Temporal Data Mining Algorithm."*

Based on these definitions, temporal data mining can be considered as mining temporal data from temporal database. Temporal data are those which are organized based on time or certain sequence order. Roddick and Spiliopoulou (2002) has determined four broad categories of temporality within data, i.e. static, sequences, timestamped, and fully temporal. Static data contains no temporal information, but the temporal inference can be made through reference to transaction-time such as referring to audit trails or transaction logs. Sequence data is an ordered list of events. The temporal information can be extracted based on the sequences. Compare to sequences, timestamped contains more temporal information as it is a timed sequence of static data taken at certain intervals. Total temporal information can be found in fully temporal category as each tuple in a time-varying relation in the database may have one or more dimensions of time.

Temporal pattern from the database can be extracted using window sliding technique. Sliding window technique is used to capture the time delay within the data set. Sliding window technique was proven able to detect patterns from temporal data (Ku-Mahamud et al., 2009; Li and Lee, 2009). This process is called segmentation process. Modeling the temporal events can be performed using artificial intelligence techniques such as NN (Moisao and Pires, 2001; Shanmugasundaram et al. 1997; Zehraoui and Bennani, 2005).

NEURAL NETWORK

Neural Network (NN) is an algorithm that has been developed based on generalizations of mathematical models of biological nervous system. It dynamically inherits human neuron information processing capability (Graupe, 1997). This capability enables NN to perform a brain like function such as forecasting, classification, and pattern matching. The NN model is used to predict the consequence of the given action. NN can be categorized into single and multi layer network. Single layer network is a model that consists of input and output layers while multi-layer network consists of at least one hidden layer between input and output layer.

In this study, standard BP NN with bias, learning rate and momentum are used to classify the rules of reservoir water release. BP NN uses a feed forward topology, supervised learning and the backpropagation learning algorithm (Bigus, 1996). The role of NN is to learn the rule pattern by creating a mapping between the input data (premise) and the target output (consequent). This mapping was established by training the NN to minimize the square error (*SE*) between the target (t) and the network output (y) (Equation 1).

$$SE = \frac{1}{2}\sum_{k=1}^{m}(t_k - y_k)^2 \quad (1)$$

Where m is the total number of output units.

RESEARCH DESIGN

Data Acquisition

In this study, Timah Tasoh reservoir was used as a case study. Timah Tasoh reservoir is one of the largest multipurpose reservoirs in northern Peninsular Malaysia. Timah Tasoh located on Sungai Korok in the state of Perlis, about 2.5km below the confluence of Sungai Timah and Sungai Tasoh. Timah Tasoh reservoir covered the area of 13.33 Km^2 with the catchment area 191.0 Km^2. Its maximum capacity is 40.0 Mm^3. Timah Tasoh reservoir serves as flood mitigation in conjunction to other purposes: water supply and recreation. Water from Timah Tasoh is used for domestic, industrial and irrigation.

In this study, a total of 3041 daily data from Jan 1999 – April 2007 were gathered from the Timah Tasoh reservoir operation record. Operation of Timah Tasoh reservoir was influenced by upstream rainfall which was manually recorded through 5 upstream gauging stations. Rainfall observed from these stations will eventually increase the reservoir water level. In this study the current water level (t), tomorrow water level ($t+1$), and the changes of water level at t, $t-1$, ..., $t-w$ were used as the input data, while the gate opening/closing at t is used as the target or the expected outcome. The constant t and w represent time and days of delays (which later represented as window size).

Data Processing

Data were imported into MS Excel and sorted based on the date. A column that represents gate opening/closing was clean to remove noise. Gate opening/closing value is in range of zero to six. Zero indicates gate is closed and values from one to six indicate the number of gates that are open. Change of this value implies the decision point. At this point window slice will be formed begin from that point and preceding to w days according the window size. In this study, the segmentation processes based on sliding window technique begin with window size 2, that represent 2 days of delay. The maximum window size was set to 10. Each segmentation process will return a total of 124 instances. Redundant and conflicting instances are then removed. Table 1 shows the usable number of instances and the window size.

Table 1. Data Set and the Number of Instances

Data Set	Window Size	Number of Instances
1	2	43
2	3	54
3	4	71
4	5	82
5	6	95
6	7	109
7	8	113
8	9	118
9	10	119

Classification Method

In this study, nine NN models were developed. Each NN model is trained with one data set. Inputs of all data sets are normalized using min-max method and rescale into a range of [-1,1]. The output was represented based on Binary-Coded-Decimal (BCD) scheme. The value "0" is replaced with "-1" so that all outputs are in the range of [-1,1]. Each model is trained with different combination of hidden unit, learning rate and momentum. Training is control by three conditions (1) maximum epoch (2) minimum error, and (3) early stopping condition. Early stopping is executed when the validation error continue to arises for several epochs (Sarle, 1995). Fig. 2 shows the pseudo code for the NN training. The aim of this procedure is to get the combination that gives the best result. Prior to the training, each data set is randomly divided into three different sets (Bigus, 1996): training (80%), validation (10%) and testing (10%) sets (Table 2).

```
for each hidden unit (HU)
where HU = {3,5,7,9,11,13,15,17,19,21,23,25}

  for each learning rate (LR)
  where LR = {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}

    for each momentum (Miu)
    where Miu = {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}

      Training:
            Feedforward()
            Backpropagation of error()
            Weight update()

      Validation()

    end loop (Miu)
  end loop (LR)
end loop (HU)
```

Figure 2. Pseudo Code for NN Training

Table 2. Division of Data Sets

Data Set	Number of Instances	Training	Validation	Testing
1	43	35	4	4
2	54	44	5	5
3	71	57	7	7
4	82	66	8	8
5	95	75	10	10
6	109	87	11	11
7	113	91	11	11
8	118	94	12	12
9	119	95	12	12

FINDINGS

The results of NN training, validation, and testing are shown in Table 3. Overall, the lowest error achieve for training, validation and testing was 0.065795, 1.59E-07, and 9E-10 respectively. The best results of training, validation, and testing was 98.35%, 100%, and 100% respectively. These results show that NN classifier has performed well on temporal rules. Based on the results in Table 3, data set 4 is chosen to be the best data set. NN train with data set 4 achieves 93.94% of training performance and 100% of validation and testing performance. The error was 0.23505, 0.023383, and 0.007085 respectively. Data set 4 was formed with window size 5 with 82 instances.

Values for the network parameters that were achieved from the training phase are shown in Table 4. As for data set 4, the total epoch is 86 and the best result achieved was with learning rate (LR) 0.8 and momentum (Mom) 0.2. The best network architecture achieved is 8-23-2.

Table 3. Results of Training, Validation and Testing

Data Set	Training		Validation		Testing	
	%	Error	%	Error	%	Error
1	90.00	0.39996	87.50	0.5	100.00	9E-10
2	90.91	0.362563	100.00	0.007216	100.00	6.13E-05
3	95.62	0.147186	85.72	0.626408	100.00	0.034537
4	93.94	0.23505	100.00	0.023383	100.00	0.007085
5	89.34	32.00295	100.00	1.59E-07	100.00	1.4E-07
6	97.70	0.092475	95.46	0.188657	100.00	0.002146
7	98.35	0.065796	100.00	0.032103	95.46	0.191186
8	93.09	0.276602	95.84	0.166669	95.84	0.168359
9	97.37	0.104647	95.84	0.171619	100.00	0.003985
Min	89.34	0.065795	85.72	1.59E-07	95.455	9E-10
Max	98.35	32.00295	100	0.626408	100	0.191186

Table 4. NN Parameters

Data Set	Epoch	#Input	#Hidden Unit	#Output Unit	LR	Mom
1	77	5	25	2	0.9	0.4
2	42	6	23	2	0.8	0.4
3	33	7	17	2	0.7	0.3
4	86	8	23	2	0.8	0.2
5	31	9	9	2	0.9	0.8
6	31	10	7	2	0.7	0.5
7	54	11	5	2	0.5	0.5
8	42	12	25	2	0.4	0.8
9	27	13	9	2	0.4	0.6

DISCUSSION

In this study, reservoir water level data which includes the current, the (expected) tomorrow water level and the changes of water level were extracted from the reservoir operation record. In actual reservoir operation and decision making, the current water level represent the current stage of reservoir water level (t), while the tomorrow water level is water level that is expected for tomorrow at $t+1$. Theoretically, this water level can be forecasted based on hydrological variables (Wan-Ishak et al., 2010). The changes of reservoir water level represent the increase or decrease of reservoir water level. Observing the changes of reservoir water level at time t and the preceding $t-1$, $t-2$, ..., $t-w$ will give an insight on when to release the reservoir water.

The sliding window technique has been successfully employed on reservoir water release data, to extract the changes of the reservoir water level that lead to the water release decision, which is opening/closing of reservoir's gate. The findings reveal that window size 5, which represent 5 days of observed water level changes contribute to the best classification performance of NN classifier. This information is vital for reservoir management to plan early water release.

The finding of this study has also shown that NN architecture 8-23-2 has produced the acceptable performance during training (93.94%), validation (100%) and testing (100%). In addition, training the network takes only 86 epochs.

CONCLUSION

Findings of this study provide and alternative information to the reservoir operator to make early decision of reservoir water release. Manually, reservoir operator monitors the changes of water level and consults the superior officer before taking the appropriate action. Having unpredicted circumstances of the weather, early decision of the reservoir water release is always a difficult decision.

Early water release from reservoir will reserve enough space for incoming inflow due to heavy upstream rainfall. In addition, water release can be controlled within the capacity of the downstream river. Thus flood risk downstream due to extreme water release from the reservoir can be reduced. In this study, window sliding has been shown to be a successful approach to model the time delays, while NN was shown as a promising modeling technique.

ACKNOWLEDGEMENT

The authors' most appreciation to the Perlis Department of Drainage and Irrigation for permission and supplying Timah Tasoh reservoir operational data.

REFERENCES

Bigus, J. P. (1996). *Data Mining with Neural Networks.* New York: McGraw-Hill

Dunham, M. H. (2002). *Data Mining: Introductory and Advance Topics.* New Jersey, Pearson Edu.

Graupe, D. (1997). *Principles of Artificial Neural Networks.* Singapore: World Scientific Publishing

Hejazi, M. I., Cai, X., and Ruddell, B. L. (2008). How Reservoirs were Operated-Exploring the Role of Hydrologic Information. *Proc. of the World Env. & Water Resources Congress*, 1-9

Jain, S. K. and Singh, V. P. (2003). Chapter 11: Reservoir Operation. In S. K. Jain and V. P. Singh (Eds), *Water Resources Systems Planning & Management, 51*, 615-679

Ku-Mahamud, K.R., Zakaria,, N., Katuk, N. and Shbier, M. (2009). Flood Pattern Detection Using Sliding Window Technique, *Third Asia Int. Conf. on Modeling & Simulation*, 45-50.

Labadie, J. W. (2004). Optimal Operation of Mutireservoir Systems: State-of-the-Art Review. *Journal of Water Resources Planning and Management, 130 (2)*, 93-111

Laxman, S., Sastry, P. S. (2006). A Survey of Temporal Data Mining. *Sadhana, 31(2)*, 173-198.

Li, H. F. & Lee, S. Y. (2009) Mining Frequent Itemsets Over Data Streams using Efficient Window Sliding Techniques. *Expert Systems with Applications, 36*, 1466–1477

Lin, W., Orgun, M. A., Williams, G. J. (2002). An Overview of Temporal Data Mining. In Simoff, S. J., Williams, G. J., Hegland, M., *Proc. of the Australian Data Mining Workshop,* 83-90

Moisao, R. L. M., Pires, F. M. (2001). Prediction Model, Based on Neural Networks, for Time Series with Origin in Chaotic Systems. *Proc. of the Workshop on Artificial Intelligence for Financial Time Series Analysis*, Univ. of Porto

Roddick, J. F., and Spiliopoulou, M. (2002). A Survey of Temporal Knowledge Discovery Paradigms and Methods. *IEEE Transactions on Knowledge and Data Engineering, 14(4)*, 750-767.

Sarle, W. (1995). Stopped Training and Other Remedies for Overfitting, *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, 352-360

Shanmugasundaram, J., Prasad, M. V. N., Gupta, A. (1997). Temporal Data Mining. *Working Paper PROFIT-97-31*, Massachusetts Institute of Technology

Smith, K. and Ward, R. (1998). *Floods: Physical Processes and Human Impacts.* England: Wiley.

Wan-Ishak, W. H., Ku-Mahamud, K. R. and Md-Norwawi, N. (2010). Reservoir Water Level Forecasting Model Using Neural Network. *Int. J. of Computational Intelligence Research, 6(4),* 947-952.

Wurbs, R. A. (1993). Reservoir-System Simulation and Optimization Models. *Journal of Water Resources Planning and Management, 119(4),* 455-472.

Zehraoui, F., and Bennani, Y. (2005). New Self-Organizing Maps for Multivariate Sequences Processing. *Int. J. of Computational Intelligence and Applications, 5(4),* 439-456.