Journal of Hydrology 365 (2009) 213–224

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Optimal tree-based release rules for real-time flood control operations on a multipurpose multireservoir system

Chih-Chiang Wei [a], Nien-Sheng Hsu [b,*]

[a] Department of Information Management, Toko University, No. 51, Sec. 2, University Rd., Pu-Tzu City, Chia-Yi County 61363, Taiwan
[b] Department of Civil Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan

ARTICLE INFO

Article history:
Received 24 November 2007
Received in revised form 20 November 2008
Accepted 25 November 2008

Keywords:
Flood control
Decision-tree algorithm
Release rules
Reservoir operation

SUMMARY

This study presents a methodology to establish a set of optimal operation release rules which are tree-based rules for real-time flood control on a multipurpose multireservoir system. The derived rules can be used to determine the optimal real-time releases during flood periods. Steps of the proposed methodology involve: (1) collection of data, (2) building of flood database, (3) generation of optimal input–output patterns by running the flood control optimization model, (4) classification of training and testing data, (5) extraction of tree-based release rules for designed scenarios using the decision-tree algorithm (C5.0), (6) determination of optimal tree-based rules, (7) generation of the real-time forecast data by using the hydrological forecast model, (8) processing of reservoir real-time releases by simulating the reservoir real-time flood control operation, and (9) verification of the superior release rules through comparisons of tree-based rules, regression-based rules derived from a multiple-linear regression model and existing release rules. The developed methodology is applied to the Tanshui River Reservoir System in Taiwan to extract the decision trees for each scenario and then select the best ones with highest accuracy as the optimal tree-based rules. The derived optimal tree-based rules, regression-based rules and existing rules are compared by conducting the real-time operations in three historical typhoons, including Aere, Haima and Nock-ten in 2004. Results demonstrate that the solution using the derived tree-based rules have better performance than the regression-based rules and the existing rules in terms of reducing the peak stage at downstream control points, and meeting the target reservoir storage at the end of flood.

Introduction

Taiwan has subtropical climate and high mountains with steep slopes all over the island. The most severe disaster in Taiwan is flooding caused by typhoons. Once a typhoon strikes, due to receiving voluminous rainfalls from upstream watersheds, considerable upstream flows usually converge downstream within a few hours. To reduce the downstream peak flow, reservoirs have been built to serve for flood control purpose during typhoon. However, the water level at the downstream may rise abnormally if the upstream reservoirs are not operated properly. With all these unfavorable conditions, it has been one of the challenges for engineers to release such voluminous water wisely from the multireservoir systems into ocean to mitigate the downstream flood disaster.

Currently, the most common release strategy for reservoir flood control operations in Taiwan involves predefined operation rules, such as the release look-up tables for flood control, where releases are expressed as a function of reservoir variables (water levels) and hydrological inputs (reservoir inflows). As an example, Water Resources Agency (WRA, 1984) stipulated flood control operation rules for the Shihmen Reservoir in order to mitigate flood damage. The flood control operation rules (called the "existing rules" in this paper) adopt the release look-up tables for each of two flood stages (i.e., peak-flow-preceding stage and the peak-flow-proceeding stage) to standardize the water releases during typhoon periods. These release tables are graded by total forecasted rainfall, the observed storage level, and the reservoir inflow during flood periods.

Although, using these predefined operation rules is straightforward, the ranges between the release look-up tables are too large to operate the release of a reservoir precisely (Chang and Chang, 2001). Another problem is that in the multiple-reservoir systems each reservoir usually has individual release rules for its operation purposes; for example, in Tanshui River Reservoir System, the two reservoirs (i.e., Shihmen and Feitsui reservoirs) stipulated the release rules separately. During extreme flood events, it has difficulty for operators in identifying the best joint release policy to mitigate downstream flood hazard.

* Corresponding author. Tel.: +886 2 33662640; fax: +886 2 33665866.
E-mail addresses: d89521007@ntu.edu.tw, nsshue@ntu.edu.tw (N.-S. Hsu).

0022-1694/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2008.11.038

To derive the operation rules for a reservoir system, the decision-tree algorithm is one of the novel technologies. The decision-tree algorithm is a powerful and popular approach for classification and prediction (Apté and Weiss, 1997; Bradford and Fortes, 2001). Decision trees extracted by the decision-tree algorithm can be used to assist decision-makers to make proper decisions. It has been widely applied to decision judgments, screening images, load forecasting, diagnosis, marketing and sales and so on (Witten and Frank, 1999). In the field of water resources management, Solomatine et al. (2000) built decision trees in classifying surge water levels in the coastal zone depending on the hydrometeorological data. Bessler et al. (2003) built the release operation rules of water supply for a multireservoir system using the decision-tree algorithm.

In order to establish the optimal operation rules of flood control during typhoon periods for a multipurpose multireservoir system, a methodology for deriving tree-based rules is developed in this article. The derived optimal tree-based rules are compared with classical regression-based rules and existing rules by conducting the real-time operations in historical typhoons. In this study, the tree-based rules are extracted by the decision-tree algorithm (i.e., induction tree technique, C5.0, the C4.5 commercial version) as described in Quinlan (1993, 1996). The developed methodology is applied to the Tanshui River Reservoir System in Taiwan.

Development of methodology

Procedures of the methodology

In this section, a methodology is presented for extracting the optimal release rules of real-time flood control for a multipurpose multireservoir system. Fig. 1 illustrates the flowchart of the proposed method. Each step in Fig. 1 can be thoroughly described as follows.

Step 1: collect data. The related data collected include reservoir data, historical flood events, and existing reservoir release rules. Reservoir data involve the capacities of the hydraulic facilities and the relationships between storage and water level. Historical flood events comprise hydrological hydrographs (including reservoir storage levels, reservoir inflows and outflows, lateral flows, water levels of downstream control points and estuary).

Step 2: build flood database. This database is built to refine the collected data (from Step 1). The data mining process requires the domain expert to perform good selection and preprocessing of data (Fayyad et al., 1996). That is to say, an important step to ensure the success of extracting release rules is to identify the selection and preprocessing steps for the data set.

Step 3: construct the flood control optimization model to obtain the optimal input–output patterns. The formulation of this deterministic model formulated as the mixed-integer linear programming (MILP) problem is described in "Flood-control optimization model". Then, the optimization model using the data from Step 2 runs to obtain the optimal release hydrographs. The optimization model is carried out by the LINGO solver (Lindo Systems Inc., 2001).

Step 4: Classify the optimal patterns into training and testing datasets. The optimal patterns obtained from Step 3 are separated into two datasets. The first dataset is used for training rules and the second dataset is used for testing the extracted tree-based rules.

Step 5: Construct the decision-tree model to extract the tree-based release rules for designed scenarios using training dataset. Two terms are defined in this study. The "attribute" (or called "variable") refers to a single data item common to all cases under consideration, and the "case" (or called "record") is the collection of attribute values of a specific case. To train a C5.0 decision-tree model, one or more In attributes (inputs) and one Out attribute

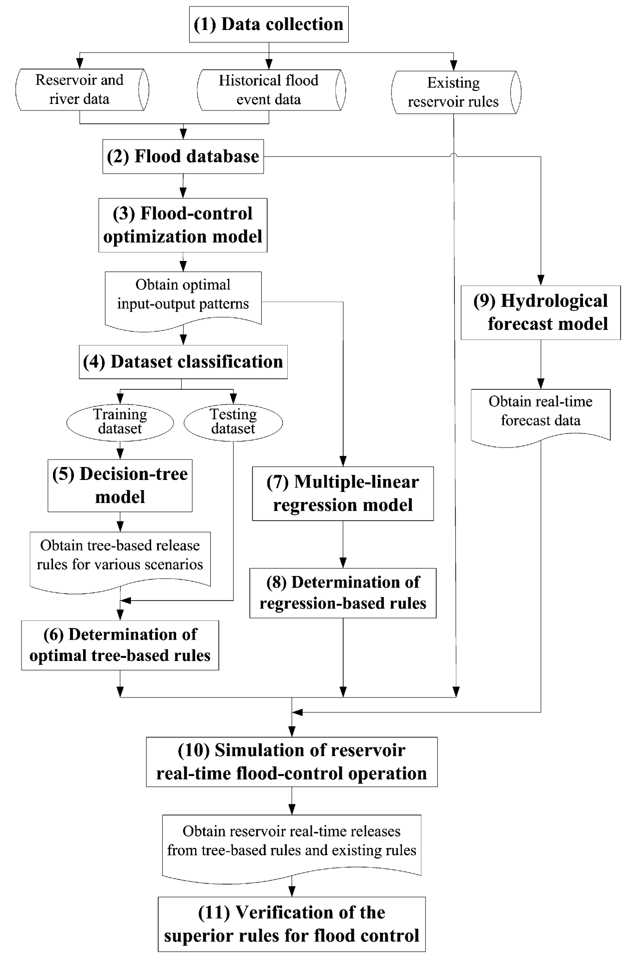

Fig. 1. Flowchart of the developed methodology for extracting optimal reservoir release rules.

(output or target) are needed. For C5.0, the In attribute can be numerical or symbolic (or categorical and nominal) type but the Out attribute must be symbolic. The principle of constructing decision trees is described in "Decision-tree model". The C5.0 algorithm is performed using the Clementine software (SPSS Inc., 2002). In this step, to extract the release rules for flood control, three sub-steps are described as follows.

5a. Find the attributes that possibly affect the behavior of reservoir releases during flood, such as water storage and reservoir inflow.

5b. Design various scenarios under the considerations of different combinations of attributes.

5c. Utilize the C5.0 algorithm to extract the graphical "tree" (the tree-based rules) for each scenario using the training dataset.

Step 6: Determine the optimal tree-based rules by using testing dataset. In order to determine the optimal rules, three sub-steps are made as follows.

6a. Define the performance criterion "Accurate" as

$$\text{Accurate } (\%) = \frac{\text{Correct trials}}{\text{Total simulation trials}} \tag{1}$$

where "Accurate" is the percentage of correct classification for total trials; "Total simulation trials" denotes the total number of simulation trials (cases); and "Correct trials" is the number of simulation