Dam Influent Prediction by Machine Learning

Nobuaki Takase,
Osamu Harakawa

Keywords Hydraulic power generation, Optimal power generation plan, Flow rate prediction, Machine learning, Ensemble learning

Abstract

In planning the daily power generation plan in a hydroelectric power plant, many constraints such as dam influent rate, dam-designated water level, and maintenance discharge should be considered. This work is extremely complicated. We developed a predictive approach for the dam influent rate by using a machine learning. Such a new technique for dam influent rate prediction is expected to be a method to establish a model for predicting future influent rates based on past influent data and meteorological data. It is used to realize a highly accurate prediction without considering physical characteristics in a river system. In the developed technique, an approach called "ensemble leaning", which is a type of machine learning, is applied. Using multiple simple and coordinative learners, the developed method realized a superb prediction accuracy even in rainy weather when obtainable data is minimal.

1 Preface

When planning a daily power generation at a hydropower plant, the dam influent rate, dam-designated water level, maintenance discharge rate, and many other restrictive conditions must be taken into consideration. This work is extremely complicated. If a future influent rate can be predicted with high accuracy, control of dam water level and discharge rate becomes easy, and stable and efficient plant operation can be carried out.

In order to achieve accurate prediction of future dam influent rate, however, it is necessary to establish a physical model with which future meteorological data, the shape of river systems, amount of snow fall, and other critical information can be accurately grasped. Such a practice is challenging. For example, although an outstanding physical model can be established for a single river system, another physical model must be established from the beginning when another river system is added to this project.

Meanwhile, if a machine learning model is made to learn the relationship among meteorological data, past influent rate data, and future influent data, a highly accurate prediction of the dam influent rate becomes possible without direct consideration of physical characteristics of river systems. This paper introduces a high-accuracy dam influent rate prediction approach by using a machine learning model.

2 Outline of Hydropower Generation Business Operations

Fig. 1 shows an outline of business operations at a dam-type hydropower plant. An example of power generating operations for 24 hours to be officially notified to a power company at 9:00 every day is presented. Essentials for the calculation of power generation plan are shown below.

(1) The dam water level is maintained within the upper and lower limits of the specified water level.
(2) For the protection of ecological systems in rivers, maintenance discharge (minimum discharge rate) is secured.
(3) The higher the dam water level, the higher the power generating efficiency. The dam water level is, therefore, kept as high as possible.
(4) When the dam water level exceeds the specified level, overflow (gate discharge to prevent dam overflow) is inevitably caused.

In order to secure high-efficiency plant operation, it is necessary to set up the water level as high as possible to maximize power generation efficiency for (3) while restrictions of (1) and (2) are satis-

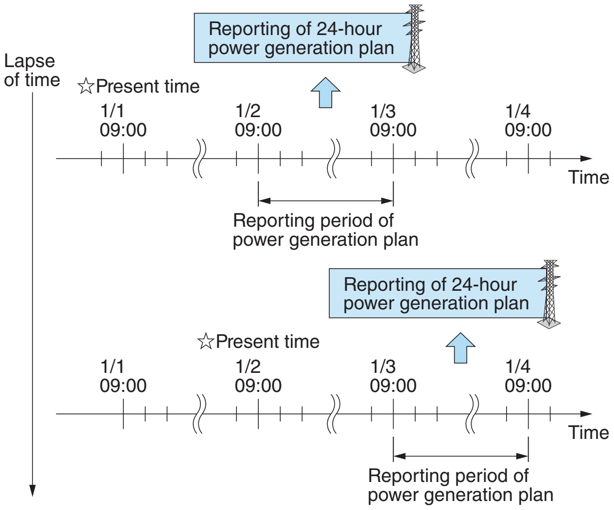

Fig. 1 Outline of Business Operations at a Dam-Type Hydropower Plant

This diagram shows an example of business operations when a daily power generation plan for 24 hours toward the next day is reported at 09:00 to the power utility company. The lapse of time goes from upper to lower. In this diagram, an example of an operation plan at a time point of 1/1 09:00 is shown in the upper diagram and a plan at a time point of 1/2 09:00 is shown in the lower diagram.

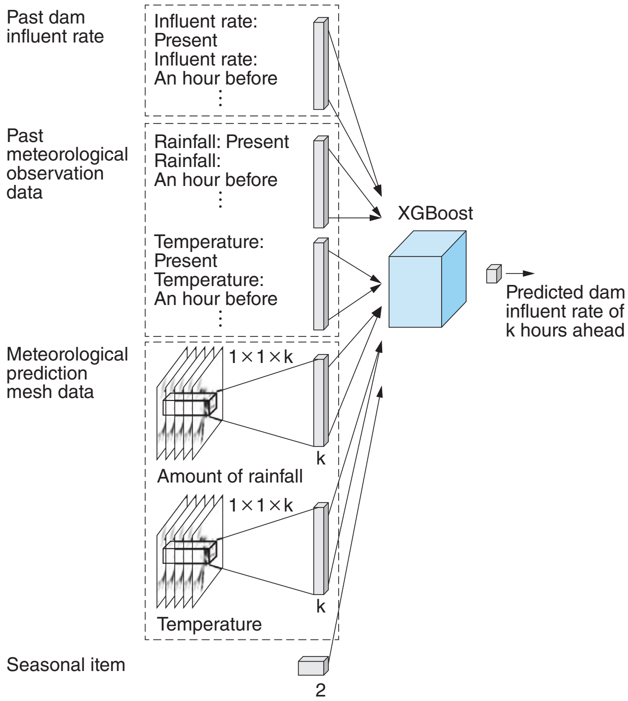

Fig. 2 Dam Influent Rate Prediction Model

When all data of past influent rates and meteorological information are entered in the XGBoost prediction model, the predicted dam influent rate of time k (= 1, ⋯ , 48) can be obtained.

fied. At the same time, overflow must be prevented for (4).

For these reasons, it is necessary to grasp future influent rate accurately in order to establish a high-efficiency power generating plan. The dam catchment area (area occupied by river systems) is, however, generally wide, and in addition, mountain ranges may be snowbound in winter. In many cases, it is therefore difficult to achieve future influent rate predictions.

3 Dam Influent Rate Prediction Model

Fig. 2 shows a dam influent rate prediction model. Input data for this prediction model generally involves actual data of past-to-present influent rates, actual meteorological data obtained in the past to present, predicted meteorological data, and seasonal items. The meteorological data input entered in the dam influent rate prediction model is considered to cover the amount of rainfall, temperatures, atmospheric pressures, and humidity. This model employs the amount of rainfall and temperatures. In addition to actual data of present dam influent rate data and actual meteorological data, it is now possible to investigate time-serial trends of each data if the past several hours of actual data input is acquired.

For meteorological prediction data, the MesoScale Model (MSM) Grid Point Value (GPV)[1] is used. For GPV, the atmosphere is sectioned by regularly aligned mesh points and forecast meteorological values are obtained at each grid point based on various observatory data by using a supercomputer. The MSM is a numerical forecast model that can manage the coverage of Japan proper and its offshore areas. Since this model can generate forecast outputs of prediction 8 times a day at the intervals of 3 hours and 39 hours ahead (51 hours ahead at 0:00 and 12:00), it is suitable for meteorological prediction at a time span of several hours to one day ahead. This development is for Japanese hydropower generation business operations. Since the purpose of prediction is for the acquisition of dam influent data for 48 hours ahead, we adopted the MSM for our numerical prediction model. It should be noted that the use of the dam influent rate prediction model from our R&D activities is not limited to Japan. It is available throughout the world through proper selection from numerical prediction models.

Factors for causing an increase in dam influent rate are rainfall and snow melting. An increase in the influent rate due to snow melting can be predicted based on the amount of snowfall in the river basin and actual temperature data gained in previous several hours. It is, however, difficult to accurately grasp actual snow amounts in river basins. For our development, we focused on snow melt occuring shortly after the snow season. As such, we devised an approach by which seasonal items are inputted in the dam influent prediction model. Fig. 3 shows an outline of seasonal items. The seasonal items are variables that fluctuate during the period of one year. By making the prediction model to definitively learn the snow-melting season and other periods, it is possible to predict an increase in dam influent rate due to the snow melt.

For the learning of dam influent rate prediction model, ensemble methods are used. For this model in particular, a development approach called the eXtreme Gradient Boosting[2] ("XGBoost" hereafter) is used.

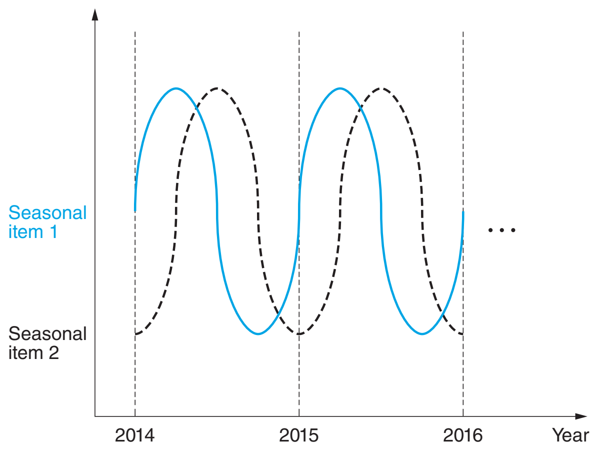

Fig. 3 Outline of Seasonal Items

The seasonable items are acceptable if they are of a periodic function of a one-year cycle. In this case, the seasonal item 1 is expressed by a cosine wave and the seasonal item 2 is expressed by a sine wave.

3.1 Ensemble Learning Method

The Ensemble Learning Method is a kind of machine learning method. In machine learning, the model learns the relationship among data groups based on large amounts of input/output data. When the unknown input data is put into the model, it outputs inference results. In our development, the input data is the past dam influent data or meteorological data while the output data is the future dam influent data. A feature of the ensemble learning method is that a high-accuracy learner can be established through coordination of multiple weak learners. Fig. 4 shows an example of a bugging approach for the ensemble learning method. In order to construct a model where regression lines are given to the data shown on the left, the following steps are executed:

(a) Split into different conditions
(b) Build Weak learners
(c) Majority/mean of the weak learners

In (a), data is split into different samples by making sampling with replacement by the bootstrap sampling method. In (b), learning is performed with a weak learner (comparatively simple learning model) for the respective samples split in (a). In learning model for a weak learner, a decision tree is generally used. In (c), output values of a final model are obtained through the processing of majority, mean, and total sum of the respective outputs from weak learners made under (b). Since

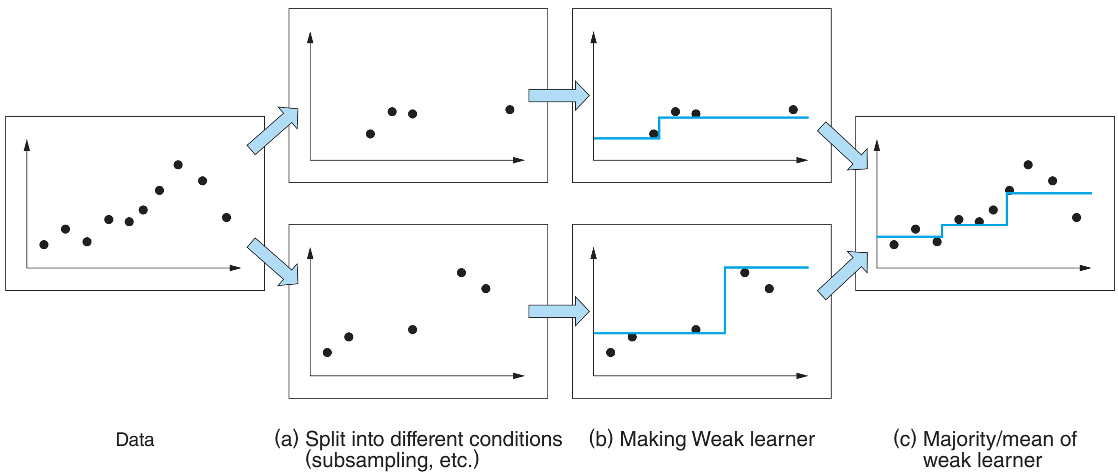

Fig. 4 Example of Bugging Approach for the Ensemble Learning Method

Black points are data. This is an example where blue lines of data regression are drawn by two weak learners.

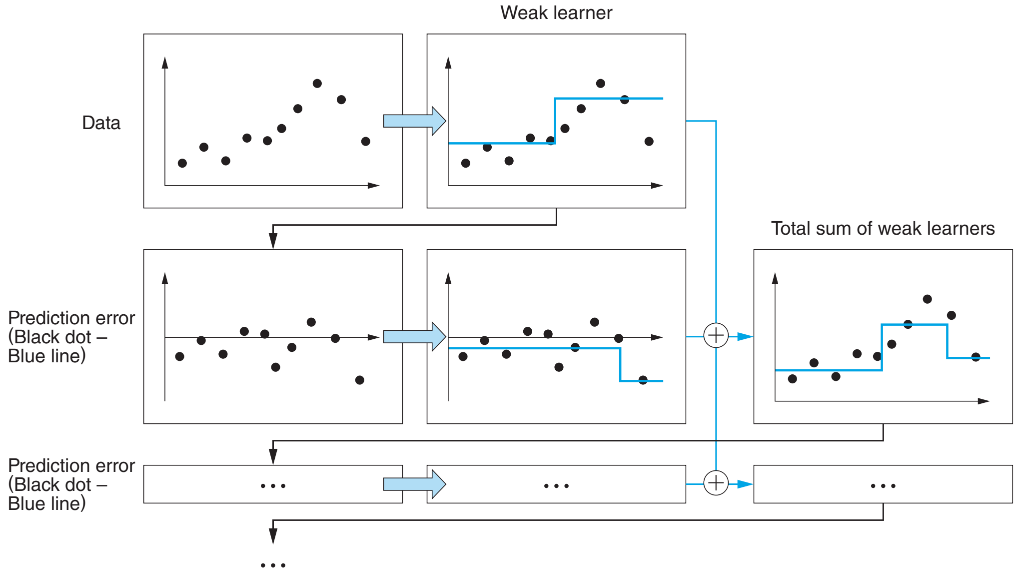

Fig. 5 **Outline of Boosting Technique**

For a prediction error caused by a weak learner, the creation of another weak learner is repeated to decrease prediction errors.

computation cost for weak learner making is generally low, ensemble methods can assure high-speed and high generalization performance.

3.2 XGBoost

In the case of XGBoost, the weak learner is made to learn with an algorithm called "boosting" which is different from bugging explained in **3.1** above. Fig. 5 shows an outline of the boosting technique. By boosting, a learning model is established by producing a new weak learner where errors generated from former weak learners are minimized. In addition to this boosting technique, the XGBoost is equipped with an algorithm where suppression of over-speed learning and overfitting is devised. For this reason, performance generalization can be obtained even for a region where errors can be easily caused and the volume of data is small. For our development, it is preferable to make it possible to accomplish precise restriction when the dam influent rate becomes high in the case of heavy rain. Data of rainy weather, however, is only about 20% of the overall data volume and such data tends to become smaller in the case of heavy rain. It is, therefore, difficult to construct a learning model that can make accurate prediction while the amount of dam influent is large at the time of heavy rain. If the XGBoost is adopted, we can expect accurate prediction even though the dam influent rate is high when the volume of data tends to be small. For our development, the XGBoost was adopted for ensemble learning.

4 Numerical Experiments

In our simulation, the effectiveness of the proposed approach for this research is shown. In this case, we made a prediction of dam influent rate relating to a dam facility in Japan. With the dam influent rate prediction model equipped with the XGBoost, we entered inputs of actual dam influent data, amount of rainfall, and actual meteorological data gathered in the previous 12 hours to the present time. For meteorological forecast mesh data by the MSM, inputs of mesh data of about 34-hours of rainfall and values of temperature forecast gathered close to the objective dam were entered. The learning period was set at 1 July 2008 to 31 December 2016. The verification period was set at 1 January 2017 to 30 November 2017.

Figs. 6 to 8 show a comparison of actual flowrate values and predicted flowrate values. Fig. 6 is for rainless days, Fig. 7 is for snow melting days, and Fig. 8 is for heavy rainy days of 17.5 mm at a moment. Fig. 9 shows actual rainfall values and predicted rainfall values by MSM measured in the same period for Fig. 8.

In Fig. 6, the influent rate on a rainless day varies from approximately 10 m^3/s to 20 m^3/s and the prediction model makes an accurate forecast. In

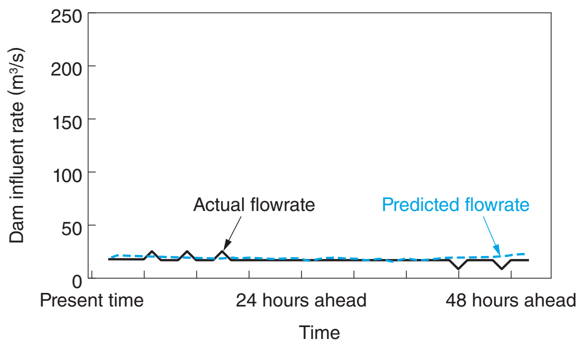

Fig. 6 **Comparison of Actual Flowrate Values and Predicted Flowrate Values: Rainless Days**

The result of prediction in rainless days is shown. The solid line indicates actual values of dam influent rate and the dotted line shows prediction output values from the dam influent prediction model.

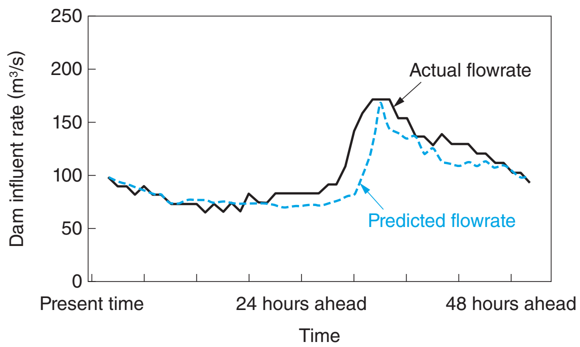

Fig. 7 **Comparison of Actual Flowrate Values and Predicted Flowrate Values: Snow Melting Days**

The result of prediction at the time of snow melting is shown. The solid line indicates actual values of dam influent rate and the dotted line shows the prediction output values from the dam influent prediction model.

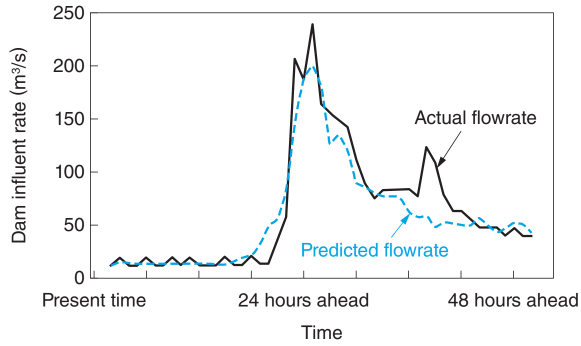

Fig. 8 **Comparison of Actual Flowrate Values and Predicted Flowrate Values: Heavy Rainy Days**

The result of prediction in heavy rainy days is shown. The solid line indicates actual values of dam influent rate and the dotted line shows prediction output values from the dam influent prediction model.

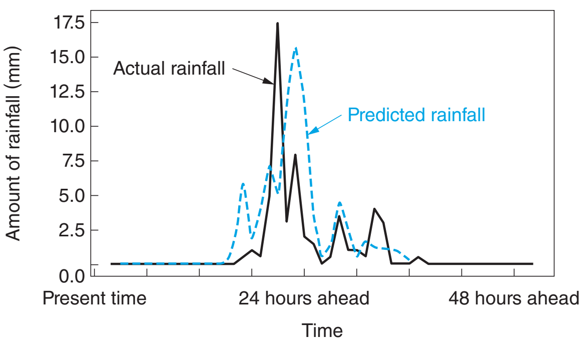

Fig. 9 **Actual Rainfall Values and Predicted Rainfall Values: Heavy Rainy Day**

Actual and predicted rainfall values at the time of heavy rain are shown. The solid line indicates actual rainfall values and the dotted line shows predicted rainfall values by the MSM.

Fig. 7, the predicted flowrate shows a good consistency even though the actual flowrate changes around more than 100 m³/s due to snow melting. Obviously, the seasonal items of the prediction model shown in Fig. 2 duly identifies the snow-melting season. The occurrence of snow melting is estimated based on meteorological data. In Fig. 8, it is possible to confirm a sufficient prediction of trends to increase the flowrate when heavy rainfall results in the rise of the influent level from 10 m³/s to 200 m³/s, though occasional errors can be seen at the time of the rising. Since the MSM achieves an accurate prediction of future rainfall as shown in Fig. 9, it is possible to consider that the learning model has successfully forecast the correct dam influent rate based on the given future rainfall data.

Table 1 shows the prediction accuracy throughout the simulation period. The mean absolute error of one-day accumulation denotes an index to show whether the total amount of influent for a single day can be predicted. The prediction accuracy is evaluated based on the predicted dam influent rate, mean

Table 1 **Prediction Accuracy throughout the Simulation Period**

The result of a day and that of the next day are shown. The mean absolute error of one-day accumulation means an index to show whether the total amount of influent for a single day can be predicted.

Evaluation index	Prediction error
Mean absolute error for 24 hours of a day measured at the intervals of an hour	10.7%
Mean absolute error for 24 hours of the next day measured at the intervals of an hour	13.6%
Mean absolute error of one-day accumulation for a day	7.0%
Mean absolute error of one-day accumulation for the next day	10.1%

absolute error of actual values, and 95 percentile※1 of actual dam influent rate, using the formula: (mean absolute error) / (95 percentile value) × 100%. Based on Table 1, we confirmed that the 24-hour forecast error of a day is about 10%, the mean absolute error of one-day accumulation of a day is 7%, and that a high-accuracy prediction is performed throughout the verification period.

The learning time of the prediction model is about 600 seconds. Compared with deep learning technology using a neural network that requires thousands to tens of thousands of seconds, we can declare that our prediction model learning by ensemble learning approach assures high-speed performance.

5 Postscript

This paper introduced a high-accuracy dam influent rate prediction approach by using a machine learning model for the purpose of optimal management of hydropower generation plants. This approach can construct a simulation model without any dependence on singular characteristics of river systems. Since the learning time for ensemble learning is short, this technology can be expected to promote a wider application to other river systems.

Regarding the data of the Mesoscale Model Grid Point Value (MSM) used for our development, we use the database[(3)] collected and distributed by Research Institute for Sustainable Humanosphere.

(Note)

※1. 95 percentile value: A dose value corresponding to the upper 5% group when the result of dose computation is placed in the order from higher to lower values.

《References》

(1) Meteorological Business Support Center: MesoScale Model Grid Point Value, http://www.jmbsc.or.jp/jp/online/file/f-online10200.html

(2) Tianqi Chen, Carlos Guestrin: "XGBoost: A scalable tree boosting system", Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785-794, ACM, 2016.

(3) Research Institute for Sustainable Humanosphere of Kyoto University: Sustainable Humanosphere database: http://database.rish.kyoto-u.ac.jp