

CNN-LSTM Coupled Model for Prediction of Waterworks Operation Data

Kerang Cao*, Hangyung Kim**, Chulhyun Hwang***, and Hoekyung Jung****

Abstract

In this paper, we propose an improved model to provide users with a better long-term prediction of waterworks operation data. The existing prediction models have been studied in various types of models such as multiple linear regression model while considering time, days and seasonal characteristics. But the existing model shows the rate of prediction for demand fluctuation and long-term prediction is insufficient. Particularly in the deep running model, the long-short-term memory (LSTM) model has been applied to predict data of water purification plant because its time series prediction is highly reliable. However, it is necessary to reflect the correlation among various related factors, and a supplementary model is needed to improve the long-term predictability. In this paper, convolutional neural network (CNN) model is introduced to select various input variables that have a necessary correlation and to improve long term prediction rate, thus increasing the prediction rate through the LSTM predictive value and the combined structure. In addition, a multiple linear regression model is applied to compile the predicted data of CNN and LSTM, which then confirms the data as the final predicted outcome.

Keywords

Big Data, CNN, Correlation Analysis, Deep-Learning, LSTM

1. Introduction

It is difficult for water purification plant operators to determine whether the data measured in real time in the field is wrong or not because it must be judged based on various data. Therefore, the need for long-term prediction that can recognize sensor state change is more crucial [1-4].

In this paper, we propose a model to improve the prediction rate of each data for stable process operation in water purification plant, and this enables prediction of demand, automatic detection of abnormal data, and prevention of accidents. In order to analyze the characteristics of the big data (water plant data), it is difficult to obtain a significant correlation coefficient among all data analysis and to investigate the suitability of the LSTM model that predicts the $t(1)$ value using $t(-n)$ to $t(0)$ data. This data was the past operational data for each item [5-8]. The prediction of the long-short-term memory (LSTM) model enables a user to obtain a stable value in the short-term prediction that predicts the

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Manuscript received October 1, 2018; first revision November 21, 2018; accepted November 27, 2018.

Corresponding Author: Hoekyung Jung (hkjung@pcu.ac.kr)

* School of Computer Science & Technology, Shenyang University of Chemical Technology, Shenyang, China (caokerang@hanmail.net)

** Jeonju Office, K-water, Daejeon, Korea (kimdwh@kwater.or.kr)

*** Consulting Dept., Data Maroo Co., Daejeon, Korea (chhwang@einssnc.com)

**** Dept. of Computer Engineering, Pai Chai University, Daejeon, Korea (hkjung@pcu.ac.kr)

value of $t(1)$, but long-term prediction is difficult to be applied for process control and abnormal value classification. Therefore, in this study, convolutional neural network (CNN) model is introduced to apply correlation values to many items and to extract multi-item correlation features like image information processing.

Moreover, we analyzed the prediction rate of LSTM model, which has the advantages of short-term prediction and the CNN model, which has multi-item correlation features. So, we propose CNN and LSTM combination model that has both the advantages of short-term prediction and high prediction rate in long-term prediction. Also, the characteristics of the combined model are set with the independent variables of the predicted value of the CNN model and the predicted value of the LSTM. We propose a multiple linear regression model which consists of control data defined as the dependent variables. Through this experiment, we compare the proposed model with the predicted value of LSTM and confirmed the performance improvement.

This paper consists of four chapters. Section 1 explains the outline of the study and Section 2 explains the proposed model. Section 3 presents the results of a comparison with the existing deep-learning model to evaluate the proposed model. Finally, Section 4 presents the conclusion.

2. Design and Implementation of CNN-LSTM Coupling Model

This section describes the design content and implementation of the CNN-LSTM combination model.

2.1 CNN_LSTM Coupling Model Design

The proposed prediction model of plant operation data solves the long-term prediction problem of LSTM. Long-term prediction is an immensely important technology for water management using water purification plant operation data. Particularly, in order to provide 'action response time' to operators and managers, one should predict data for a relatively long time because the $t+1$ data provided by the LSTM is insufficient.

The proposed model combines CNN, which is mainly used for image classification, with LSTM for time series prediction. This collaboration raises the accuracy of long-term prediction of LSTM.

The main considerations of the proposed model are as follows:

- Long-term predictions will be affected by the state or environment of the plant operation where the data are extracted. In other words, if a similar environment can be found in the past operations, that data can be utilized to improve prediction performance in a positive way.
- A clustered image can be created by making a two-dimensional matrix using multiple sensor data input at a certain time. The created image will be similar to the storage format that is often used for image processing, and the CNN method can be applied.
- CNN can find data similar to the input of two-dimensional matrix from past data. Because CNN has an ability to classify missing images through the convolution, some of the data used in LSTM can be used to find historical data needed for long-term prediction.
- Past similar data sets extracted through CNN and combined with long-term predictive values of LSTM will possess more accurate values.

In this paper, the proposed model is composed of LSTM, which is a deep learning [5] method, mainly used to predict time series of data for water purification plant operation data and CNN, which is mainly used for image classification. The method of combining the data derived from these two deep learning models uses a multiple linear regression model.

The main idea of this paper is to deduce the past situation similar to the time when the prediction is needed by using the classification ability of CNN and to calculate the predicted value by combining the data at certain time and the data predicted by LSTM. This further maximizes the benefits of each model.

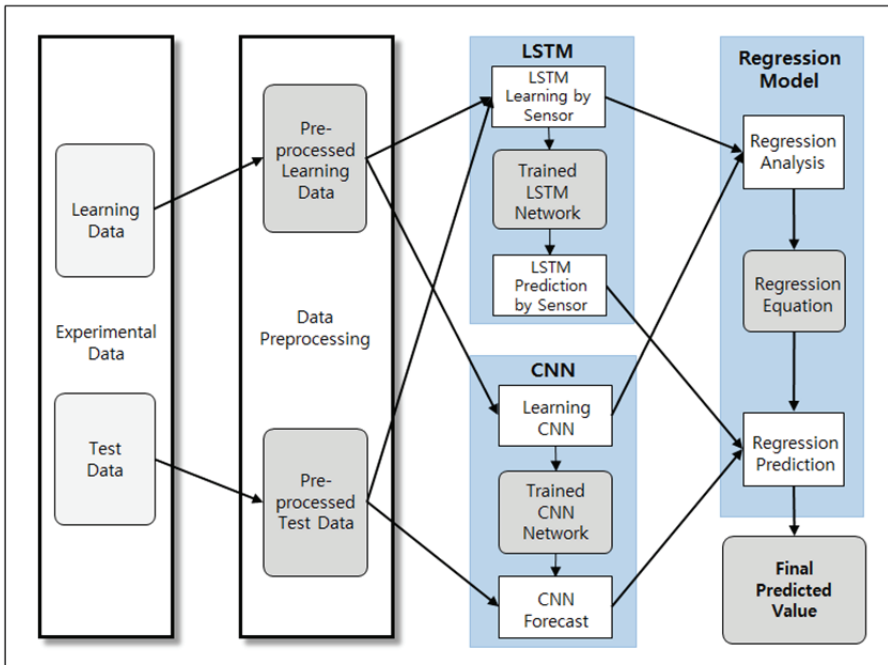


Fig. 1. Conceptual construction of proposal model.

Fig. 1 shows the preprocessing, learning and prediction processes that make up the proposed model.

The flow of the proposed model is either learned or predicted through ‘data preprocessing → LSTM → CNN → regression model’. However, the two steps of LSTM and CNN model can be performed in any chosen order. This is because the two models receive the same inputs and generate different outputs and then combine into the regression model. The following describes the process configuration, definitions, and roles for each step.

- **Learning Data:** Among the experimental data, we use pre-classified data of the deep learning neural network learning presented in the proposed model. In this paper, the data of the last three months with an exclusion of the test data are extracted to be used as learning data.
- **Test Data:** Among the experimental data, certain data is classified in advance to test the prediction performance of the proposed model. In this paper, we extracted recent 7-day data among all data and used it as test data.
- **Data Preprocessing:** The process of converting input sensor data into a form that can be used in a predictive model.

- The missing value generated from the input data is replaced with an arbitrary value (0 in the experimental environment) and divided into a matrix form in a way that chosen data can be processed by the CNN model. After partitioning, we assign a cluster according to the similarity of each matrix. Classification of clusters assigned to each image is essential in the processing of CNN, which has the characteristics of a supervised learning.
- Pre-processed Learning Data: Data is transformed into learning data that can be processed by a deep learning neural network model. Pre-processed learning data has the following three characteristics. The first characteristic is that a missing value is removed to be used in both LSTM and CNN. The second characteristic is a matrix form of $n \times m$ to facilitate processing in CNN (Transformed into a matrix of $n \times m$ for processing in CNN which is optimized for image classification. In that matrix, n is the sensor and m is the time). In other words, it is the data that divides the data of the whole sensor by m time. The third characteristic shows the feature classification value of each matrix composed of $n \times m$. The purpose of extracting matrix feature classification values is to perform CNN supervised learning.
- Pre-processed Test Data: This data represents the prediction data converted into the form that can be processed by the deep learning neural network model. Preprocessed predict data is the result of removing missing values and converting those values into substitute value. This conversion allows the data to be processed in the predict model first. For the CNN classification, we construct the predicted input data matrix of $n \times m$, where the first $n \times (m/2)$ part consists of the input data for prediction, and the remaining $n \times (m/2)$ part consists of the alternate (default) value.
- Learning LSTM: This system performs the prediction by inputting learning data to LSTM neural network. In order to adjust the weight of neural network, the system compares the calculated value with the actual value. Since the neural network configuration of LSTM is built up for each sensor, learning related to the neural network is also performed for each sensor.
- Learning CNN: We learn CNN neural network using matrix data and cluster result provided from preprocessing process. Learning results are adjusted according to the weighted values in CNN neural networks.
- LSTM Forecasting: We receive input test data and calculate the LSTM predicted value. Based on the data of a certain time, the system performs time-series prediction from the forecast time to $t+1 \dots t+n$ and performs prediction independently for each sensor.
- CNN Forecasting: This inputs the test data and calculates the CNN classification value. Only 50% of the learning data type is used since the system uses only the time data calculated in the LSTM prediction process. This is a way to classify the damaged image. The predicted $t+1 \dots t+n$ data is replaced with the default value.
- Trained LSTM Network: It is the data which records adjusted LSTM weight information through the learning process. The stored learning results are used as information to constitute the prediction model.
- Trained CNN Network: It is the data which records CNN weight information through learning process. The stored learning result is the information that constitutes the predictive model.
- Regression Analysis: It calculates multiple linear regression equations through regression analysis using learning data. Regression analysis uses the actual values (learning data) of CNN and LSTM.
- Regression Equation: It is data in which the variables of each term in multi-linear regression

equation are derived from regression analysis. It is the data that records the model information, which calculates the final predicted value when performing the prediction using the proposed model.

- Regression Prediction: It calculates the final predicted value by inputting CNN predicted value and LSTM predicted value into regression equation.

2.2 Implementation of the Proposed Model

The predefined main environment variables for LSTM implementation are as follows. The layers and parameters of the LSTM are configured as default values in the referenced code for comparison experiments.

- The network layer representing the neural network layer of LSTM consists of 7 layers in total. It sets the batch size to 512 which determines how many predictions to be performed after the prediction during learning process.
- Activation function, an input function of neural network, uses 'tanh', the most widely used mathematical function in time series data. The tanh function both resizes and shifts the size and position of a sigmoid function. This is traditionally used in neural networks.
- The number of epochs that determines the repetition level of learning is set to five. This value is the same as the previous research environment that reflects the loss rate of the network.

For the sensor data prediction at the time $t+1$, the proposed program implementation result uses 50 times of data from $t-49$ to $t-0$ as input data and proposes a program to predict the data of 50 times (50 minutes in the experiment) from $t+1$ to $t+50$.

50 minutes of data are presented to construct a total of 100 sensors and a square matrix (50 minutes input +50 minutes prediction), excluding the sensors that continuously generate meaningless signals. This exclusion facilitates the experiment, leaving no special restrictions.

CNN is a deep learning technology that is used mainly for images and video labels. In order to apply CNN, which is primarily used for the classified images or moving images, to the prediction model of sensor data, we apply two methods: composition of input data and label generation.

First, the model converts the collected data into a form that can be processed by CNN. The general CNN processing method has the form of $n \times m \times c$, where n is the vertical pixel of the image, m is the horizontal pixel of the image, and c is the color (Red/Green/Blue) it implies. In this paper, the sensor (n) input at the same time and the time (m) processed at one time are made into a matrix and configured in the same way as the CNN model processes.

Second, label information for supervised learning is generated. Because CNN classification is based on supervised learning classification results should be generated in advance.

In this paper, since the data to be applied to CNN is the time series sensor data of $n \times m$, the data represents measurement environment of the plant operation data which is referred by the sensor data value.

The CNN configuration information constructed in this paper is as follows. Deep learning neural networks are constructed to suit the basic network structure without additional modification to pass the verification of the proposed method. The layers and parameters of CNN are configured as default values in the referenced code for image classification for comparison experiments.

- The network layer representing the neural network layer of CNN consists of three layers in total.
- The system sets the batch size to 512 which determines how many predictions to be performed after the prediction during learning process is made.
- Activation function, which is an input function of neural network, uses ReLU (rectified linear unit) and softmax which are the most widely used functions in image classification. The ReLU and softmax functions are frequently and actively used in classification problems.
- The number of epochs that determines the repetition level of learning data is set to 20. This value takes into account the loss rate of the network.

The next step after obtaining learning data is the classification prediction process using test data. In order to perform CNN classification-prediction, time-series data must be converted into image data in the same way as the learning process works. When converting test data for prediction into an image, the system uses only $n \times (m/2)$ data, which is composed of test data. This is different from the obtaining learning data process, where only $n \times (m/2)$ data is composed of test data instead of $n \times m$ data, and time after the remaining $t+(m/2)$ is replaced with a value of '0'. The reason for replacing the data generated at $t+(m/2)$ from time $t+m$ with '0' is to generate the missing data arbitrarily since this interval must be predicted. This whole process is similar to the steps of mapping a damaged image into the most similar category in CNN learned from the original. Since the time-series data is predicted based on the similarity between the past data and the matrix configuration, a method of directly extracting the eigenvalues of the test data and determining the cluster should be taken into account. However, since 50% of the damaged data patterns exist, and the most effective classification method for damaged information is CNN, CNN is applied. But the spatial characteristics of CNN are appropriately considered. In the matrix of $n \times m$, the spatial characteristic is reflected as the column structure represents the adjacent time. However, adjacent rows are meaningless as the row configuration did not consider any characteristics of each sensor.

In this paper, a multi-linear regression model is used to integrate the predicted data from LSTM and CNN into final predicted data. Equation 1 shows multi-linear regression model.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

where Y is final prediction value; X_n , n^{th} predict value predicted value (LSTM, CNN, real value at $t-1$); and β_n , n^{th} regression coefficient value.

First, the significance of the regression model is determined through the F-statistic after the generation of the multiple linear regression models in the learning phase. The significance of the regression model is determined depending on whether LSTM and CNN prediction data and actual data can be described in linear form in the learning data. Fig. 2 shows the process in the multiple linear regression analysis.

In this paper, only regression equations are applied to the values of sensors with significance emphasized on the multiple linear regression models. The predictive error rate for the predicted value, which is the root-mean-square error (RMSE), was calculated. Eq. (1) shows a method of obtaining RMSE by comparing predicted values with real values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - R_i)^2}{n}} \quad (2)$$

Where P_i is i^{th} prediction value and R_i is i^{th} real value.

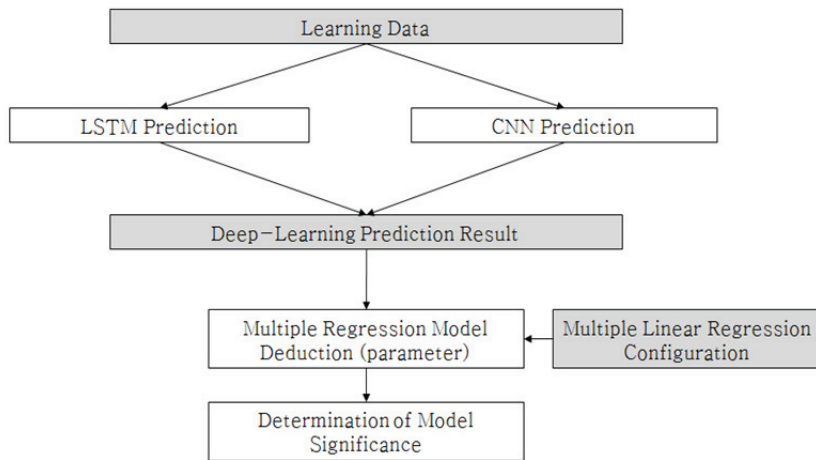


Fig. 2. Process of calculating and verifying significance of multiple linear regression model.

In order to compare the performance of the prediction model in the paper, we use the RMSE, which is the method of calculating the error by comparing each predicted value with the actual value. The smaller the value calculated from the mean square root error (closer to 0) means the better prediction performance.

3. Evaluation of Proposed Model

In this chapter, the evaluation criteria and the results are presented to compare the prediction performance of the proposed prediction model with the existing methods.

3.1 Evaluation Criteria of the Proposed Model

This section presents evaluation criteria to verify whether the proposed model shows better performance than the existing model in terms of making long-term predictions. For the evaluation of two or more prediction models, the following criteria are set and applied to the evaluation.

- First, the error rate calculation method utilizes the RMSE used in the accuracy evaluation.
- Second, the accuracy criterion does not take a concept of precision into account because the predicted data is a numerical value, thus not a search or classification. The criterion only considers how much the predicted value is closer to the actual value.
- Third, the initial configuration of the LSTM used in the proposed model is the same as that of the existing method. And CNN network is also configured the same as when CNN alone used the initial configuration for prediction. In other words, all models that make up the combined model have the same DNN state as when each model was used separately.
- Fourth, the experimental data used for the evaluation also has the same data type and value as each model is independently performed.

- Fifth, these criteria for network status and data are pure methods for measuring the ‘*Ripple effect of combine*’.

3.2 Experimental Environment

For estimating the model accuracy of water purification plant operation data, the hardware used was windows 10 Pro, CPU is Intel Core i5-7400, RAM 8 GB, software is Python 3.5.2 and deep learning is Tensorflow 1.7.0, Keras 2.1.6. The experimental environment was performed at the personal computer level. To overcome various constraints in the learning and prediction stages, I/O was managed at the file level, not the DBMS level.

The data applied to the experiment is the actual data extracted from the field, which manages the water treatment plant operation data. The environment related to the data is as follows.

- Learning Data Set: The accumulation period of learning data is one year (132,480 pieces) with data recorded every 1 minute. In addition, the number of sensors used in the learning data is 100, excluding the sensor that includes a large number of wrong-missing data.
- Test Data Set: The accumulation period of learning data is 7 days (10,080) set in every 1 minute. The number of sensors used in the learning data is 100, excluding the sensor that contains a large number of wrong-missing data.
- From the test data, the data for the prediction experiment was extracted from 10 parts of the data set of 100 minutes derived from the 7-day data, allowing the experiment to be repeated over 10 times.

3.3 Experimental Procedure

The experimental procedure to evaluate the prediction performance of the proposed model is as follows.

This section describes the experimental procedures and methods of the learning phase.

- LSTM Learning: LSTM network is learned by using experimental data. Internally, learning data and verification data are classified and used again. The difference between the learning data and the verification data occurs depending on whether it is used purely for learning or for adjusting the state of the network. The ratio of the data used as learning and verification in the experiment is 7:3. The result of the learning of LSTM comes from the learned LSTM network.
- Matrix Construction: The matrix configuration is to transform the data structure into a form suitable for performing CNN classification. In the experiment, a matrix was created in 100 (minute units to have a matrix structure of 100×100). The matrix data is generated every 100 minutes so that data is not duplicated due to the limitation of the experimental system.
- Eigen Value Calculation: In CNN, the eigenvalues are calculated from the previously extracted matrix. Since the eigenvalue has one value for each row, it is calculated as a vector of 1×100. The process of repeatedly obtaining the eigenvalue leads to a target label data.
- CNN Learning and Verification: CNN is learned by using the extracted matrix and the eigenvalues. The label l value, which is the classification information for the matrix, is used by extracting two digits of the eigenvalue. Learning and validation data are used in a 7:3 ratios as in LSTM.
- Multiple Linear Regression Analysis: The multiple linear regression analysis is performed using the predicted values of each model derived from the verification process. The calculated multiple

linear regression models are performed using the combining method (combination formula) of each model in the prediction step.

In this section, the experimental procedure and method of the prediction step are explained.

- **CNN Classification:** CNN classification is performed using 50 minutes data (100×50) immediately before the target (100 sensors, 50 minutes) are predicted. The classification result is provided as label information (eigenvalue) of the learned data set.
- **CNN Prediction Data Extraction:** A data set having the same label as the label information extracted in the CNN classification step is extracted as prediction data from the learning data set. Since the prediction data is not classified into an equal number, the number of prediction data sets differs at each prediction.
- **LSTM Prediction:** The predicted data of the target period is calculated through the LSTM. Since the prediction is performed for each LSTM derived from each sensor, the prediction takes the form of a one-dimensional matrix (1×50). The total predicted matrix (100×50) is constructed through a combination of all of the predicted matrices for each sensor.

In this section, we describe the experimental procedure and method of the test phase for comparing and evaluating the results of performing each of model predictions.

- **RMSE Calculation:** The average error is calculated by comparing the prediction result of each model calculated in the previous step with the actual value. The average error is the difference between each predicted value and the actual value divided by the actual value.
- **RMSE Comparison:** By comparing the mean error values calculated in each model, we compare the RMSE between the proposed combination model and the single model.

The test data set for the experiment carries out learning and prediction experiments based on a total of 100 sensor data except for the sensor which contains a large number of wrong-missing data.

Each image presented in Fig. 3 shows various time-series characteristics for the experimental data of each sensor. The types of sensors that generate water purification plant operation data demonstrate various aspects (such as flow meter, water level, pipe pressure, turbidity, pH, etc.), meter and valve opening, pump motor condition, water supply status and facility status. The Y axis is the size unit value for each sensor, while the X axis is the unit time in 1 minute. Among the various 100 sensor data patterns, only two sensors such as flow rate and water level are shown.

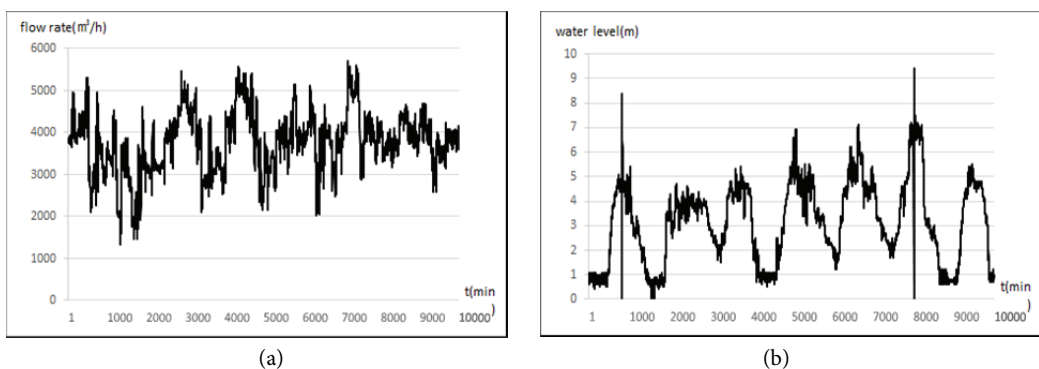


Fig. 3. Pattern of operational data sensor. (a) Flow rate and (b) water level.

3.4 Experimental Result

As a result, the RMSE calculated by the LSTM alone is significantly different depending on each sensor. It can be seen that the variation of the size of experimental data of each sensor is diverse, and numerous time-series characteristics are presented. The sensor with the smallest error is 0, and the sensor with the highest error shows an average of 2,133.

CNN resulting stand-alone experiment for prediction shows that the RMSE is very different for each sensor, and that there exists the range where the RMSE is not estimated.

The first is that the RMSE differs from sensor to sensor in the same way that the LSTM is used alone.

The second characteristic where the RMSE is blank shows that the predicted value cannot be presented because the significance level is not found in the multiple linear regression models. In other words, the CNN model extracts multiple predictive data from several historical data and does not calculate the regression equation that can combine the predictive data. In this case, the RMSE is excluded from the calculation.

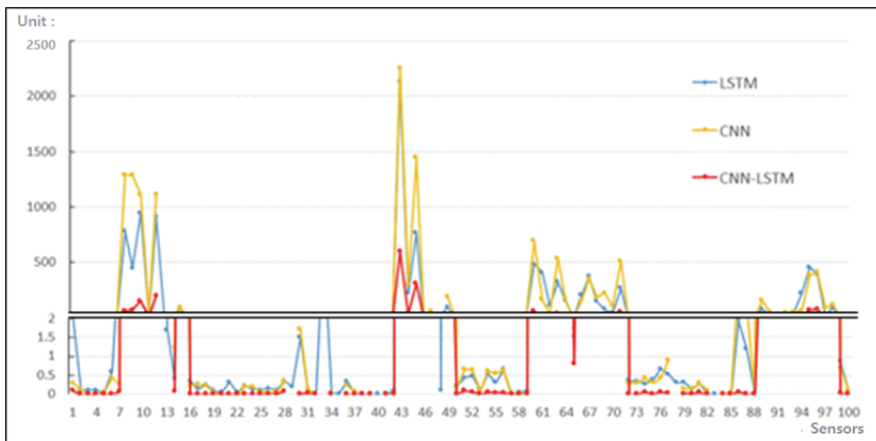


Fig. 4. The RMSE per sensor in each model prediction.

The experiment shows the prediction in the combined model in this paper, Fig. 4 shows that the RMSE is zero in the graph. This outcome occurs as in the case of the CNN alone model, no model exceeding the significance level was found in the multiple linear regression analysis.

Fig. 5 compares the results of long-term predictions of three models of LSTM alone, CNN alone, missing complement model, and CNN-LSTM combined model. As shown in Table 1, on the average, the error of the CNN single model (=132.1) is the largest in the total order (experimental type) with a value 10 times bigger. And then the LSTM alone model (=104.7) and finally the CNN-LSTM combination model (=16.4) showed the lower errors.

In particular, the combined model of CNN-LSTM shows a significantly lower error than the single model. Therefore, the proposed CNN-LSTM combining algorithm has the highest prediction performance.

This means that the performance of CNN-LSTM combined long-term prediction model proposed in this paper has a higher performance than the long-term prediction model that uses LSTM alone. It can be suggested that the long-term prediction performance of the proposed method “CNN-LSTM combined model” is higher than that of the existing LSTM-based prediction in all cases of long-term prediction.

Also, in this paper, we propose a CNN-LSTM combining model based on the assumption that CNN will support the prediction of LSTM by reproducing the previous sensor environment with only missing test data. The predicted performance of the proposed CNN-LSTM combined model is as high as 16.4, representing the statistically meaningful experimental results. Also, it can be seen that the variance of the mean error is the smallest value in the coupled model, so it has stable prediction accuracy in the experiment. This proves that it is reasonable to assume that CNN will complement the disadvantages of long-term prediction of LSTM.

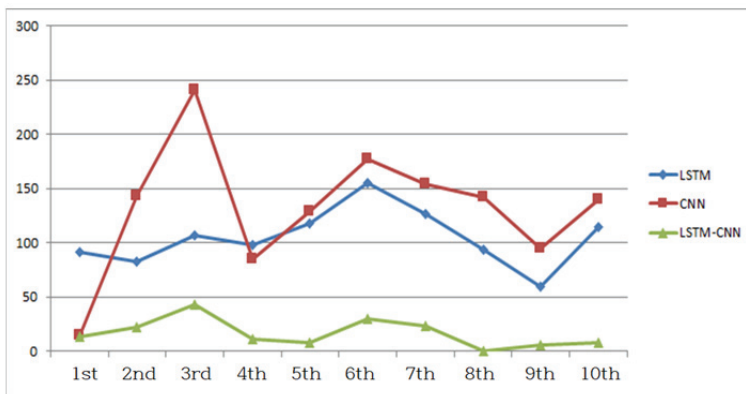


Fig. 5. Comparison chart of RMSE between models.

Table 1. Performance comparison between proposed method and existing method

	RMSE	
	Average value	Variance value
LSTM	104.7	690.8
CNN	132.1	3576.6
LSTM-CNN	16.4	166.4

4. Conclusion

The existing prediction models have been studied in various types of models such as time-series characteristics, day-series characteristics, and multiple linear regression models that contain seasonal characteristics. However, the existing model shows that the predict rate for demand fluctuation and long-term prediction are not highly accurate.

In particular, the LSTM model among the deep learning models has been applied for predicting the monitoring data of the water purification plant because of its excellent time-series prediction characteristics. However, a model that reflects various correlating aspects with an improved long-term predictability is needed.

Therefore, in this study, in order to predict the classification of each data set, CNN prediction model and improved version of the existing LSTM predictive model was employed, combining model constructions as well as applying multiple linear regression models. Also, the improved model enables the user to synthesize the predicted data of LSTM and CNN and confirm data as the final predict data.

As a result, the LSTM single RMSE is about 104.7 in average among 10 experiments. And the RMSE applied by the CNN model alone is 132.1, which showed that a LSTM has a better RMSE than the CNN model. The CNN-LSTM combined model proposed in this study has an RMSE of 16.4, which is better than the existing prediction model.

In the future, it will be necessary to improve the calculation method for many of the input variables applied to CNN and LSTM models in order to improve learning and prediction time. In the operation of water purification plant, the performance index that can process short-term and long-term prediction of many input variables in real time should be applied to increase the field applicability of the model proposed in this paper.

Acknowledgement

This study was supported by the research grant of Pai Chai University in 2018.

References

- [1] J. R. Kim, G. W. Shin, S. T. Hong, and G. H. Yoo, "A study on detecting algorithm for outlier data in water supply," in *Proceedings of the Korean Institute of Communications and Information Sciences Summer Conference*, 2016, pp. 327-328, 2016.
- [2] J. H. Seok, J. J. Kim, J. Y. Lee, J. J. Lee, Y. J. Song, G. W. Shin, "Abnormal data refinement and error percentage correction methods for effective short-term hourly water demand forecasting," *International Journal of Control, Automation and Systems*, vol. 12, no. 6, pp 1245-1256, 2014.
- [3] G. H. Yoo, J. R. Kim, and G. W. Shin, "A study on the cleansing of water data using LSTM algorithm," in *Proceedings of the Korea Institute of Information and Communication Engineering Fall Conference*, 2017, pp. 501-503, 2017.
- [4] H. S. Kim, "Data cleansing algorithm for reducing outlier of water supply information," Ph.D. dissertation, Pai Chai University, Daejeon, Korea, 2018.
- [5] E. J. Lee "Basic and applied research of CNN and RNN," *Broadcasting and Media Magazine*, vol. 22, no. 1, pp. 87-95, 2017.
- [6] Y. H. Kim, Y. K. Hwang, T. G. Kang, and K. M. Jung, "LSTM language model based korean sentence generation," *The Journal of Korean Institute of Communications and Information Sciences*, vol. 41, no. 5, pp. 592-601, 2016.
- [7] G. D. Sim and J. H. Seok, "A study on the prediction interest rate using LSTM systems," in *Proceedings of the Korean Institute of Communications and Information Sciences Winter Conference*, 2018, pp. 308-309.
- [8] C. Kim and J. W. Kang, "LSTM based anomaly detection on semiconductor manufacturing data," in *Proceedings of the Korea Software Conference 2017*, Busan, Korea, 2017, pp. 760-762.



Kerang Cao <https://orcid.org/0000-0002-4053-7166>

He received a B.S. degree in 2006 from Computer Science and Application of North Eastern University, China and Ph.D. degree in 2011 from the Department of Computer Engineering of Pai Chai University. From 2016 till present, he worked for Shenyang University of Chemical Technology as a lecturer. His current research interests include multimedia document architecture modeling, information processing, IoT, big data, and embedded system.



Hangyung Kim <https://orcid.org/0000-0003-1543-9350>

He received a master's degree from the Department of Electrical Engineering at Chungbuk National University in 2007 and a doctor's degree from the Department of Computer Engineering at Pai Chai University, Korea in 2018. He has worked for the Korea Water Resources Corporation since 1989. His current research interests include multimedia structure modeling and information processing.



Chulhyun Hwang <https://orcid.org/0000-0002-6265-1314>

He received an M.S. degree in 1995 and Ph.D. degree in 2017 from the Department of Computer Engineering of Pai Chai University, Korea. From 1991 to 2000, he worked for Korea Navy as a Computer Officer. Since 2001, he has worked in a variety of IT professional companies, where he now works as a data consultant. His current research interests include data architecture, master data, big data, and deep learning.



Hoekyung Jung <https://orcid.org/0000-0002-7607-1126>

He received a B.S. degree in 1987 and Ph.D. degree in 1993 from the Department of Computer Engineering of Kwangwoon University, Korea. From 1994 to 2005, he worked for ETRI as a researcher. Since 1994, he has worked in the Department of Computer Engineering at Pai Chai University, where he now works as a professor. His current research interests include multimedia document architecture modeling, machine learning, IoT, Bigdata, and AI.