Development and verification of usage prediction service for intelligent smart water service

Youngjoon Choi
Convergence Technology Research Institute
AIBLab Inc.
Seongnam, South Korea
yjchoi@aiblab.co.kr

Sungsik Park
Convergence Technology Research Institute
AIBLab Inc.
Seongnam, South Korea
sungsikp@aiblab.co.kr

Junwook Lee
Convergence Technology Research Institute
AIBLab Inc.
Seongnam, South Korea
junux@aiblab.co.kr

Yongwoo Kim
Convergence Technology Research Institute
AIBLab Inc.
Seongnam, South Korea
ywkim@aiblab.co.kr

***Abstract*— This paper proposes a real-time consumption prediction model that reflects the characteristics of consumers and time-series characteristics so that water supply can be effectively supplied by using a time-series prediction model. In order to reflect the characteristics of consumers, consumers were classified by industry, and each consumer's real-time usage was used to reflect time-series characteristics. There are two major water supply real-time usage prediction models: a short-term forecasting model that predicts daily usage and a long-term forecasting model that predicts weekly usage. Through such research and demonstration, it is expected that water supply and production plans based on predictions will be established to provide high-quality water supply and to enable efficient operation and management of water supply facilities.**

Keywords—smart water, time-series data, water usage prediction

I. Introduction

In order to supply high-quality clean water to citizens, smooth operation of water purification plants is necessary. This is because if the amount of water used by consumers is not accurately predicted and production is increased, the produced water may not be used for a long time. Therefore, it can be said that accurately determining the optimal amount of tap water through water demand forecasting plays a very important role in terms of production, supply and operation. Smart water service is a service technology that supports intelligence, efficiency, and optimization by applying ICT technology to the entire process from water production, supply, and consumption.

Existing water supply usage predictions were mainly made for a single consumer [1], block-level prediction [2], or water purification plant-level prediction [3]. This means that researches have been mainly conducted to predict the amount of water supply produced in the water supply system and the amount of consumption in terms of the consumer who ultimately consumes it. However, although the prediction method for a single consumer has high accuracy, it has a disadvantage in that real-time prediction is required for all consumers in order to use it for water supply planning. In addition, the prediction of the water purification plant unit is as accurate as the prediction of the total usage, but the prediction result has a limitation in that it cannot be used to establish a water supply plan after production. With the development of IoT technology, remote meter reading for various infrastructures in the city is being actively applied. In particular, the use of communication technologies based on LoRa or NB-IoT is increasing.

Therefore, in this study, we propose an intelligent smart water consumption prediction service that collects meter reading data through the IoT communication network linked to the digital water meter and uses it to predict real-time water consumption for each consumer unit. In order to predict real-time usage of waterworks, a real-time waterworks usage prediction model was developed by reflecting the characteristics of each customer's industry, and its effectiveness was verified by applying it to actual services.

II. Related works

In the field of remote meter reading, such as water supply and energy, research to predict consumption has been actively conducted. In the early days, statistics-based methods were mainly used, but currently, studies on usage prediction using machine learning, which can design and model relationships between variables in the learning process, are being conducted. As a recent study, [4] analyzed the collected power consumption data of apartment houses and predicted consumption patterns using Deep Neural Network (DNN). In [5], Recurrent Neural Network (RNN), Long-Short Term Memory (LSTM)) and Gated Recurrent Unit (GRU) were used to predict energy consumption in France and some parts of the United States. The study in [6] can be said to be the most similar to this study by predicting the energy consumption of individual households using an LSTM-based model.

However, in the use of waterworks, various usage patterns are shown depending on the business area, weather, time, and household information. In particular, it is necessary to present a highly accurate prediction model by generating and predicting usage models by reflecting the characteristics of industry types such as household, commercial, and public facilities for each customer.

III. SMART WATER SERVICE ARCHITECTURE

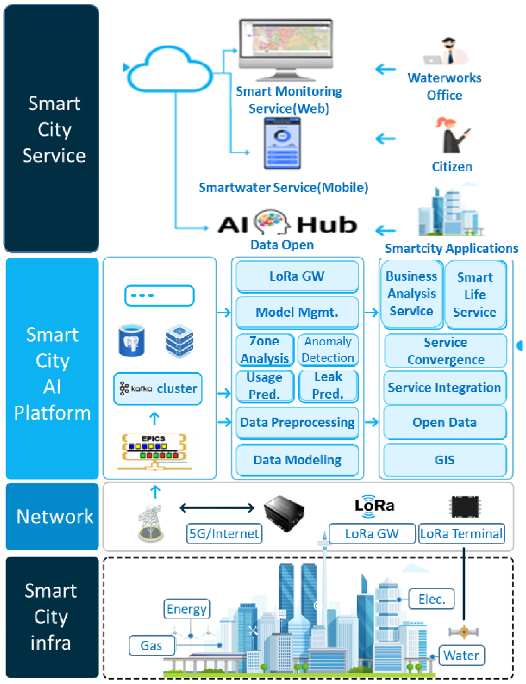

Fig. 1. Smartwater service architecture

This study was conducted as part of the smart water service. The smart water service collects water consumption data related to people's lives in real time and uses it to provide convenience to people's lives by predicting usage / progress, diagnosing commercial districts, diagnosing abnormalities of people living alone, and diagnosing water leaks. As a platform that provides , it is shown in Fig. 1. Among them, this study aims to reliably provide prediction results of usage and progressive stages to consumers who have installed remote meter readers.

IV. USAGE PREDICTION MODEL

In this study, we designed a short-term prediction model that predicts water consumption per day for each customer and uses it for operation at the reservoir stage, and a long-term prediction model that can respond to water supply plans by predicting water consumption for a week.

A. Data Preprocessing

Before implementing the short-term and long-term prediction models, in the pre-processing process, scaling of usage data and processing of outlier data caused by communication errors were carried out. Data scaling was performed to improve stability and convergence speed in the process of comparative analysis of multi-dimensional values and model optimization. In the case of usage data, it may frequently occur that the usage is higher than the existing usage due to the occurrence of an event for each customer or weather. Because the amount of usage on that day could become an outlier and affect the entire data, RobustScaler was used to perform scaling to minimize the effect of outliers.

In addition, preprocessing was performed for outliers and missing values that may occur in the collection stage of usage data. As for outliers and missing values that may occur in the collection stage of tap water usage data, a total of three types can appear, such as negative usage or missing usage, as shown in Fig. 2. The reason for the negative usage is the reverse flow of water during the measurement process, and for the case where the usage is omitted, the error in the communication network can be cited as the cause. As a countermeasure for this, the average value reflecting the history of past meter readings was used for learning and verification by replacing outliers and missing values. In the case of zero usage, the actual consumer vacated the house and there was no usage or the usage was impossible to measure, so a method of deleting and excluding customers who showed usage below a certain standard was used.

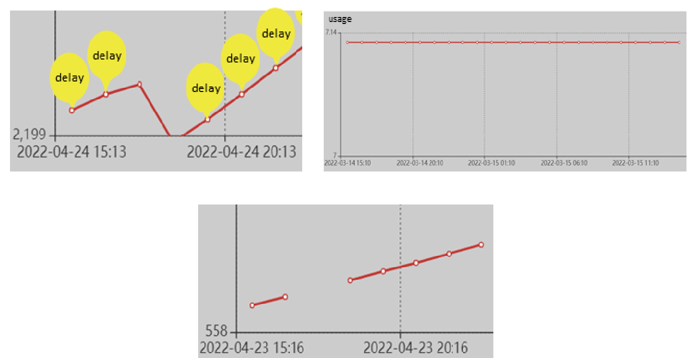

Fig. 2. Example of usage data outlier: minus value(left-top), zero value(right-top), missing value(bottom)

B. Short-Term Usage Forecast Model

In the case of the short-term forecasting model, it is a model that uses data from the past 3 days as input data and predicts the amount of usage 1 day later. To this end, the cumulative water consumption collected by each consumer was resampled on a daily basis. In addition, in the preprocessing process, customers whose usage could not be measured for more than 3 days due to poor communication or absenteeism were excluded from learning.

C. Long-Term Usage Forecast Model

In the case of long-term forecasting, the usage data resampled by 1 day in the short-term usage prediction model is resampled to the sum for 1 week, used as input data, and used as input data, and is a model that predicts usage 7 days later. The long-term usage prediction model was also not used for learning in the case of consumers who missed usage measurement for more than 7 days in a row during the preprocessing process.

Identify applicable funding agency here. If none, delete this text box.

D. *Model Implementation*

An LSTM model was used as a usage prediction model. An LSTM network is a type of recurrent neural network, and refers to a network that uses LSTM blocks for the hidden layer [7]. It is one of the neural networks designed to overcome the gradient loss problem of the existing Recurrent Neural Network (RNN), and in this study, the LSTM model was finally adopted and research was conducted. In addition, word embedding and multi-input models were constructed for the customer number in order to use the unique customer number of each customer as a weight in the LSTM model. Finally, the prediction process of the usage prediction model is shown in Fig. 3.

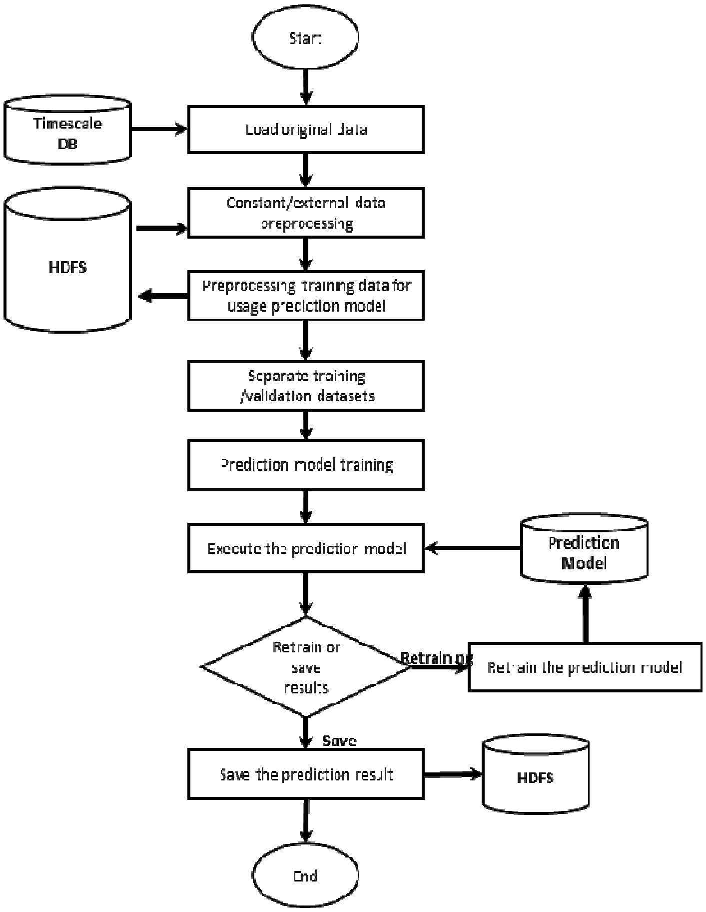

Fig. 3. Usage forecasting model forecasting process

V. Performance Evaluation and Demonstration

A. *Verification Environment*

In relation to this study, in 2022, in Dong-gu, Daejeon Metropolitan City, 1616 remote meter reading data collection infrastructures were established through LoRa private networks. Each customer collects meter reading data every hour by grouping them into 6-hour units, and meter reading data is stored in the data collection platform and provided for learning and service. Among them, a total of 1015 households were able to learn and predict through preprocessing, and 1007 households for home and general use, excluding special industries, were used for learning and verification.

Fig. 4. Digital meter(left) and meter reading terminal(right)

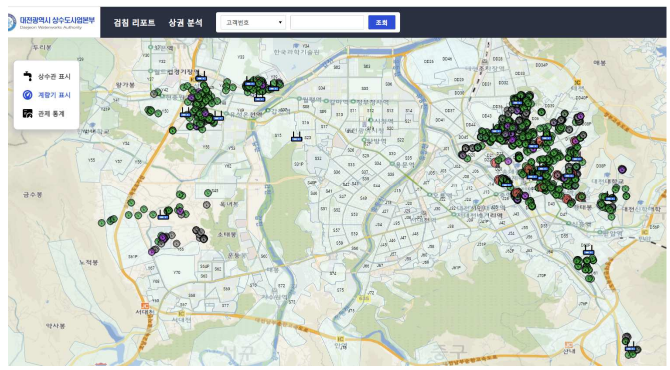

Fig. 5. Meter reading network monitoring

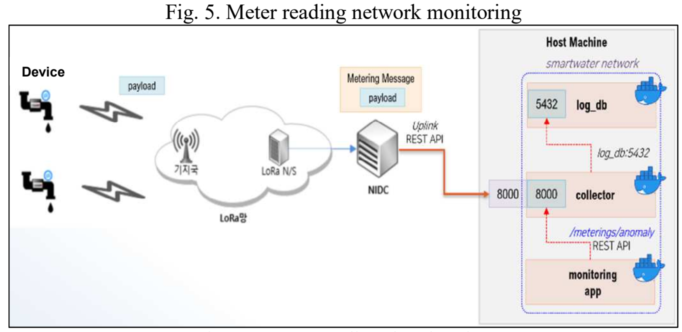

Fig. 6. Meter reading data logging

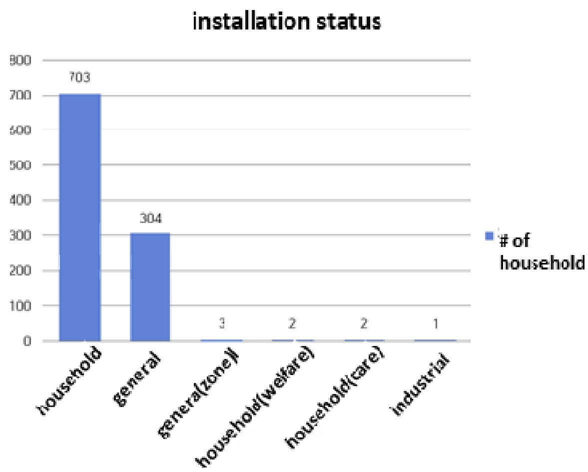

Fig. 7. Meter reader installation status by customer type

B. *Performance Evaluation*

For learning and verification, the amount of water supply used by each consumer from January 2022 to July 2022 was collected and used, which corresponds to about 40,000 cases. Model evaluation criteria were evaluated by applying MAPE (Mean Absolute Percentage Error).

TABLE I. SHORT-TERM CONSUMPTION PREDICTION MODEL RESULTS

Activation	Optimizer	Loss Function	MAPE
Relu	ADAM	MSE	8.98
Tanh	NADAM	MSE	9.05
Tanh	ADAM	MAE	**7.38**

TABLE II. LONG-TERM CONSUMPTION PREDICTION MODEL RESULTS

Activation	Optimizer	Loss Function	MAPE
`	ADAM	MSE	7.16
Tanh	NADAM	MSE	5.89
Tanh	ADAM	MAE	**3.09**

Tables 1 and 2 are tables showing the results of predicting water consumption. In the case of the short-term water supply usage prediction model, an error of 7.38% was found when using tanh as the activation function, ADAM as the optimizer, and MAE as the loss function. When using MAE as , an error of 3.09% was shown. Through this, it can be determined that it is necessary to predict real-time water supply usage by considering the characteristics of consumers and time-series characteristics.

C. *Service Verification*

Through this study, consumers were able to check the predicted amount of usage within the AI-based smart water service app one day and one week later. If existing consumers inquired through the web service to check usage/progression level, through this study and project, it became possible to conveniently view life convenience information such as current usage, predicted usage, and progress on mobile when necessary.

VI. CONCLUSION

With predictions for a single consumer or total usage because they used data produced by focusing on waterworks production. In addition, since it was not possible to collect data of individual consumers, the existing monthly meter reading data was transformed and replaced with real-time data to conduct research. However, this study was conducted because it was possible to collect the real-time consumption of each consumer through a meter reader, and it was experimentally proven that water consumption by consumer could be predicted in real time using the customer's characteristic information, weather, and time-series characteristics. In the future, research will be conducted to advance model advancement and learning by combining not only waterworks but also other energy data.

ACKNOWLEDGMENT

This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport(Grant: RS-2022-00155803).

REFERENCES

[1] Jithish, J., et al. "A neuro-fuzzy approach for domestic water consumption prediction", IEEE Region 10 Symposium (TENSYMP), 2017.

[2] Park, Sung Shik, et al. " Development of a water supply consumption prediction method considering water supply block characteristics and time series characteristics", 2020 International Conference on Interdisciplinary Research on Computer Science, Psychology, and Education (ICICPE),2020

[3] Kwon, Hyun-Han, et al. "A Development of Water Demand Forecasting Model Based on Wavelet Transform and Support Vector Machine," JKWRA Vol. 45, No. 11, pp. 1187～1199, 1997.

[4] Namjo Yoo, et al. "Analysis of Apartment Power Consumption and Forecast of Power Consumption Based on Deep Learning, .Korean.electr.electron.eng.Vol.23,No.4,1373～1380,December 2019

[5] G. M. U. Din et al., "Short term power load forecasting using Deep Neural Networks", 2017 International Conference on Computing, Networking and Communications (ICNC), pp. 594-598, Santa Clara, CA, Jan 2017.

[6] W. Kong, Z. Y. et al., "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network" , IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841-851, Jan 2019.

[7] S. Hochreiter et al , "Long short-term memory",Neural computation, vol. 9, no. 8, pp. 1735-1780, Nov 1997.

Fig. 8. Smart water service app UI