JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 22, No. 4, pp. 705-710, Apr. 2021

Check for updates

빅데이터 및 뉴럴 네트워크 기반 수질관리 미들웨어 설계 및 구현

박 성 훈[1] · 유 수 창[2*] · 방 승 범[3]
[1]충북대학교 컴퓨터공학과 교수
[2,3]충북대학교 컴퓨터공학과 박사과정

Design and Implementation of Water Quality Management Middleware based on Big Data and Neural Networks

Sung-Hoon Park[1] · Su-Chung Yoo[2*] · Seung-Peom Pang[3]
[1]Professor, Department of Computer Engineering, Chungbuk National University
[2,3]*Doctoral Course, Department of Computer Engineering, Chungbuk National University

[요 약]

최근 도시화, 산업화로 수질 오염사고 발생 피해 규모가 대형화될 우려가 커짐에 따라 사회 물 안전망에 대한 수요가 증가하고 있다. 최근 5년간 4대강 유역에서 많은 수질 오염 사고 발생하였으며, 이로 인해 용수공급 중단 및 오염된 수돗물 음용 등 국민 건강에 직간접적인 피해를 야기하였다. 따라서, 수질 환경의 불확실성을 최소화할 수 있는 수질 환경 관리 시스템이 꾸준히 요구되어 왔다. 수질관리 시스템은 Ubiquitous Sensor Network 환경에서 단말 노드에서 실시간으로 계측 된 데이터를 서버에 전송하고 이를 전송받은 시스템의 미들웨어는 데이터의 무결성과 중복성을 확보하여 저장하여 왔다. 그러나 이런 처리 과정에 가장 큰 문제는 생성된 원천 데이터는 많은 오류가 포함되어 있어 이를 그대로 사용 할 수 없다. 따라서 수질 데이터의 무결성을 확보하는 데 전문가의 조력이 필요하며 경제적으로 돈을 지불해야 하는 어려움이 있다. 이러한 문제를 해결 할 수 있는 방안으로 기계 학습 모델을 이용한 다층 신경망 구축이 최상의 해결 방안이 된다. 본 연구에서는 다층 신경망 기반 실시간 수질 데이터 검증 미들웨어를 설계하고 이를 모니터링 시스템으로 개발을 제안한다.

[Abstract]

As there is a growing concern that the scale of damage caused by water pollution accidents due to urbanization and industrialization is increasing, the demand for water safety in society is increasing. In the last five years, many water pollution accidents have occurred in the four major river basins, and this has caused direct and indirect damage to public health, such as stopping water supply and drinking contaminated tap water. Therefore, there has been a constant demand for a water quality environment management system that can minimize the uncertainty of the water quality environment. In the Ubiquitous Sensor Network environment, the water quality management system transmits data measured in real time from the terminal node to the server, and the middleware of the received system has secured and stored data integrity and redundancy. However, the biggest problem in this processing is that the generated source data contains many errors and cannot be used as it is. Therefore, the assistance of experts is required to ensure the integrity of the water quality data, and there is a difficulty in paying money economically. As a solution to this problem, building a multilayer neural network using a machine learning model is the best solution. In this study, we design a real-time water quality data verification middleware based on multi-layer neural networks and propose to develop it as a monitoring system.

색인어 : 수질관리 시스템, 센서네트워크, 빅데이터, 미들웨어, 뉴럴 네트워크

Key word : Water Quality Management System, Sensor Network, Big Data, Middleware, Neural Network

http://dx.doi.org/10.9728/dcs.2021.22.4.705

Received 18 March 2021; **Revised** 12 April 2021
Accepted 12 April 2021

***Corresponding Author; Su-Chang Yoo**
Tel: ※ 개인정보 표시제한
E-mail: izibt@nate.com

Ⅰ. Introduction

The importance of the environmental monitoring system is increasing day by day as the size of the damage increases year by year due to the rapid increase in the frequency of hail, typhoons, collapse of incisions, and landslides due to heavy rain and heavy snow caused by climate change and El Niño in Korea. In Korea, the average annual damage caused by natural disasters such as typhoons and floods has increased by 50 times from 30 billion won in the past 1970s to more than 1 trillion won. The number of rescue operations by emergency rescue teams due to natural disasters such as droughts, typhoons, hail, etc. also increased by almost four times in three years (612 cases in 2007, 554 cases in 2008, 1,313 cases in 2009, 2,347 cases in 2010). Due to urbanization and industrialization, there is a growing concern that the scale of damage will increase in the event of a water pollution accident and the demand for a social water safety is increasing. In the last 5 years, 259 cases of water pollution (Han River 99, Nakdong River 31, Geum River 25, Seomjin River and Yeongsan River 19, and 85 others) have occurred in the four major river basins. This caused direct and indirect damage to public health, such as stopping water supply and drinking contaminated tap water [1].

Therefore, it is required to establish a water quality environment management strategy system based on big data that can minimize the uncertainty of the water quality environment by expanding the target of water quality management from the current water quality management system centered on the four major rivers to small and medium-sized rivers, tributaries/branches, and reservoirs. [2].

The water quality management system transmits the data measured in real time from the terminal node to the server in the Ubiquitous Sensor Network environment, and the middleware of the received system has secured data and stored data integrity after eliminating redundancy.

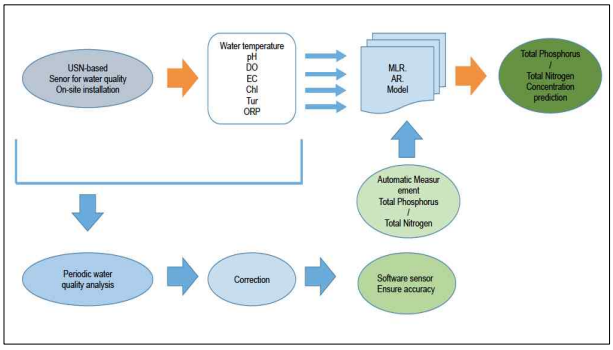

그림. 1. 수질관리 예측 센서 미들웨어
Fig. 1. Water pollution prediction software sensor middleware

Various communication methods have been studied to efficiently transmit measured data in terminal nodes, and data has been accumulated in a server in real time. In general, a USN-based water quality monitoring system needs not only a network that senses data and transmits it, but also a middleware function that collects, stores, and analyzes the sensed data and helps to have decision making based on it [2,3].

However, as the most basic problem, the assistance of water experts is required to ensure the integrity of water quality data, and there is a financial difficulty in paying money. This is because the source data that was created immediately contains many errors and many problems, such as using it as it is and showing incorrect results for use in analysis as big data. Therefore, quality control of water data has been performed by executing correction and correction work on the sensed source data with the help of experts [3,4].

It is believed that a lot of manpower and money are required to continuously modify and supplement the quality of basic water data that is sensed and collected in real time, and middleware that can improve this is desperately needed. However, in verifying data accuracy, it is difficult to verify whether the quality of water data is normal or error because environmental factors affecting water quality are complex and change dynamically over time. Because there is no mathematically standardized rule to decide this, the integrity of the water data is determined according to the situational judgment of the expert [5]. Therefore, as a solution to this problem, building a multi-layer neural network using a machine learning model is the best solution. In this study, we design a real-time water quality verification middleware system based on a multi-layer neural network and propose a development of it as a water quality analyzing system.

Big data is becoming an essential requirement for discovering and solving national social issues in fields such as healthcare, energy, climate, disaster, and economic fields. The importance and value of big data is gradually increasing, and the current utilization is expected to create a new data ecosystem by gradually spreading from the field of science and technology to all fields of public, business, and service. McKinsey estimates that the healthcare, public administration, retail, manufactur -ing and personal information sectors will also generate economic impacts of between $100 billion and $700 billion in each sector [1].

This study aims to construct a big data based middleware system that can present useful water environment information by analyzing the water quality information accumulated over a long period of time for the processing of such water quality information on big data.

Ⅱ. Related research

Figure 2 Using the multi-layer neural network in Figure 2, we design and implement middleware for verifying the accuracy of the quality of water data based on big data and multi-layer neural networks, and implement APIs for various applications based on this middleware. Figure 2 shows the information processing step based on big data and multi-layer neural networks in implementing application components such as verification of quality in environmental water data. First, the artificial neural network is trained as shown in Figure 2 through the standardized learning data. The artificial neural network is divided into three layers, and it is constructed so that the neural network can be trained by consisting of an input node layer, an output node layer, and a hidden node layer between them [6,7].

In order to construct an artificial intelligence model, an appropriate learning procedure must be followed, and an iterative process is required to adjust the number of nodes in the hidden layer and the number of learning times. Depending on the type and characteristics of the data, the appropriate number of nodes may vary, and when simulating natural phenomena that have nonlinear relationships unlike linear relationships, excessive learning can cause large errors in model validation for unlearned ideas [8,9].

However, so far, not many studies have been conducted on the number of nodes and the number of learning related to natural water phenomena, and in most cases, definitive conclusions have not been obtained. Therefore, in this paper, we try to determine the structure of a model suitable for the problem in this middleware through the learned data [10,11].

In this paper, the number of nodes and the number of learning of the hidden layer are changed to apply to verification of unlearned events to find a structure with the least error, and the verification process will be automated to avoid manual repetition. Figure 3 shows the automated parameter calculation process of the artificial intelligence model.

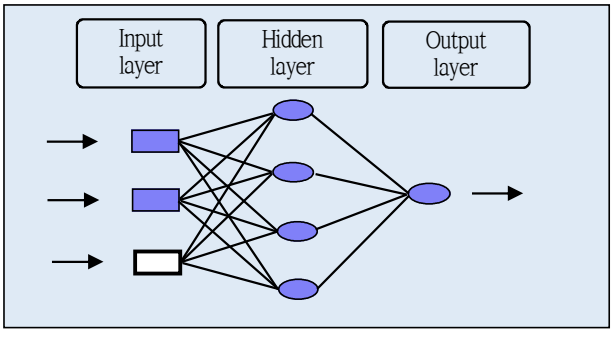

그림 2. 다층 신경망 모델의 구조
Fig. 2. Structure of a multi-layer neural network model

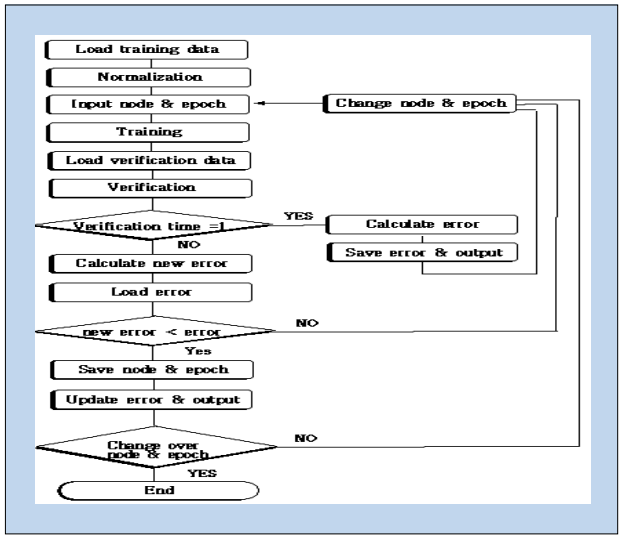

그림 3. 인공지능 모델의 자동화계산 프로세스
Fig. 3. Automated parameter calculation process of artificial intelligence model

Ⅲ. Research method and System Design

3-1 Contents, methods and results of research and development

Until now, there are standards for water quality management, but since they are not standardized, there have been problems in distortion and reliability of water data quality. Therefore, this study aims to design and build a analyzing middleware through the development of an algorithm to verify the reliability of water data quality based on multi-layer neural networks and big data to overcome this problem [12,13].

There are LSTM Model and ARIMA Model as multilayer neural network algorithms required for analyzing middleware design and construction. Using these two models, a new model for water data quality verification is designed, and the learning information of the past data and the test information of the current data are compared, and implemented as a program to meet the reliability and quality control standards of the water data. Figure 4 below shows the LSTM Network's memory block, and Figure 5 shows the ARIMA Model.

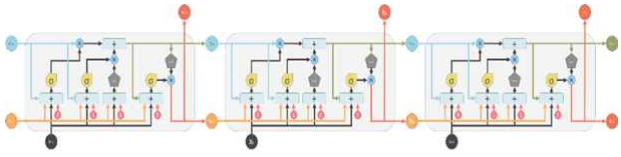

그림 4. LSTM 네트워크 모델
Fig. 4. LSTM Network Model

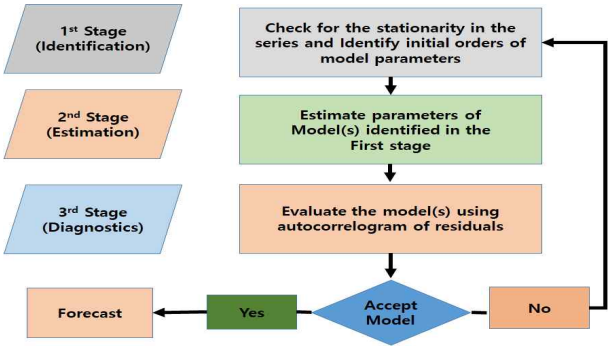

그림 5. ARIMA 모델
Fig. 5. ARIMA Model

In this paper, we implemented a multi-layer neural network algorithm for water quality verification and designed a new model, the WQ-NN (Water Quality Neural-Network) Model, based on the above two models.

The data used for training and testing were provided in the form of a CSV file, but normalization and pre-processing of the data were required, and a cleansing operation was performed on the abnormal data within the standard error range of the quality control. Figure 6 shows the water quality data set (example) after pre- treatment, and Table 1 shows the measurement unit (example) for each parameter.

Figure 7 is a schematic diagram of the detailed flow of step by step data processing. The original data is read, the preprocessing process and necessary data are extracted, and a new model is designed for the verification and evaluation of the data using the LSTM Model (CNN) and the ARIMA Model (ANN), and the designed model includes an algorithm for verification.

date		temperature	dissolved_ oxygen	ph	turbidity
0	2020-10-01	21.350596	6.898147	7.009447	13.468674
1	2020-10-02	21.315316	6.993022	7.044741	10.618085
2	2020-10-03	21.884984	6.716186	6.972112	40.292612
3	2020-10-04	20.344586	6.903656	6.928980	25.303738
4	2020-10-05	17.923966	7.364901	6.980445	26.565599
5	2020-10-06	17.950552	7.529336	7.019258	28.172014
6	2020-10-07	18.914272	7.345588	7.031518	28.425536
7	2020-10-08	20.003922	7.162172	7.031793	27.986241
8	2020-10-09	21.093359	6.888860	7.013097	30.546341
9	2020-10-10	21.421380	6.762587	7.007799	29.287538

그림 6. 전처리 후 수질 데이터 세트
Fig. 6. Water quality data set after pretreatment

Parameters	Measurement Units
Temperature	℃
pH	No units
Dissolved Oxygen (DO)	mg/L
Turbidity	FNU

표 1. 수질파라미터 측정
Table 1. Measurement units by parameter

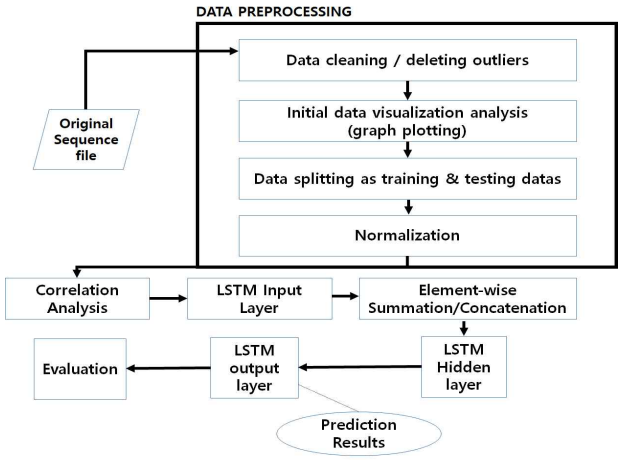

그림 7. 제안한 모형의 구조도
Fig. 7. Architecture diagram of the proposal model

Ⅳ. Experiment and results

4-1 Development tools and uses

- PyCharm: Python IDE for development tool
- Jupyter Notebook: Coding result and test tool
- Pandas: Analysis and Visualization Library
- Numpy: n-Dimensional Array
- Matplotlib: 2-D Plotting Library
- Scikit-learn: ML Library
- Keras: Neural Network Library

4-2 Simulation anf Verification

Figure 8 shows the results of learning the past data and comparing the results of the current data with the developed algorithm to compare the current data with the past data. As a result of the comparison, the measured results of the learned past data (Train) and the newly developed test data were excellent in reliability, and the measurement result was within the tolerance range of the quality control standard for outlier data and the quality control standard for the measurement result in the pre-processing process.

```
nn_model = Sequential()
nn_model.add(Dense(10, input_dim=1, activation='relu'))
nn_model.add(Dense(1))
nn_model.compile(loss='mean_squared_error', optimizer='adam')
early_stop = EarlyStopping(monitor='loss', patience=2, verbose=1)
history = nn_model.fit(X_train, y_train, epochs=100, batch_size=1, verbose=1, callbacks=[early_stop], shuffle=Fals
e)
```

```
Epoch 1/100
1339/1339 [==============================] - 2s 2ms/step - loss: 0.0361
Epoch 2/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0322
Epoch 3/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0316
Epoch 4/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0312
Epoch 5/100
1339/1339 [==============================] - 2s 2ms/step - loss: 0.0310
Epoch 6/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0310
Epoch 7/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0309
Epoch 8/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0309
Epoch 9/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0309
Epoch 10/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0309
Epoch 11/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0309
Epoch 12/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0309
Epoch 13/100
1339/1339 [==============================] - 2s 1ms/step - loss: 0.0309
Epoch 00013: early stopping
```

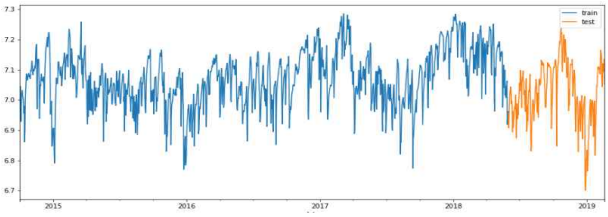

그림 8. 테스트 데이터의 측정 결과
Fig. 8. Measurement result of Train data and Test data

4-3 Experiment results

In order to confirm the results of R&D, the dataset was divided into Train and Test Dataset, and 80% was included in the Train Dataset, but 20% was set as the raw data in the Test Dataset. MixMax normalization was used to convert the values of the Train and Test Datasets in the range of -1 and 1. Figure 9 shows the basic linear graphs for the four parameters required for water quality prediction. In the line graph of Ph, it can be seen that it is in the range of 6.7 ~ 7.3. This shows that the pH of the water quality is in the ideal range. (Ideal range: 6.5 ~ 7.5)

It can be seen that the pH and DO in Figure 10 only change according to the season, but the measurement results for pH and DO show a trend without change.

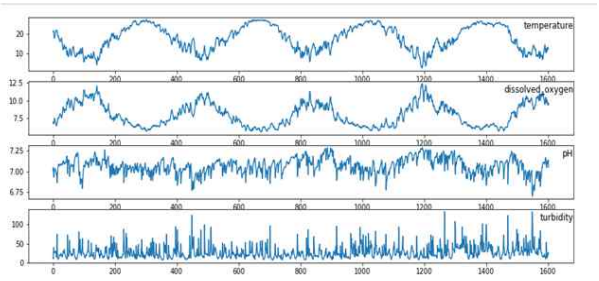

그림 9. 4개의 데이터 셋 의 추의
Fig. 9. Trend by 4 parameters of Dataset

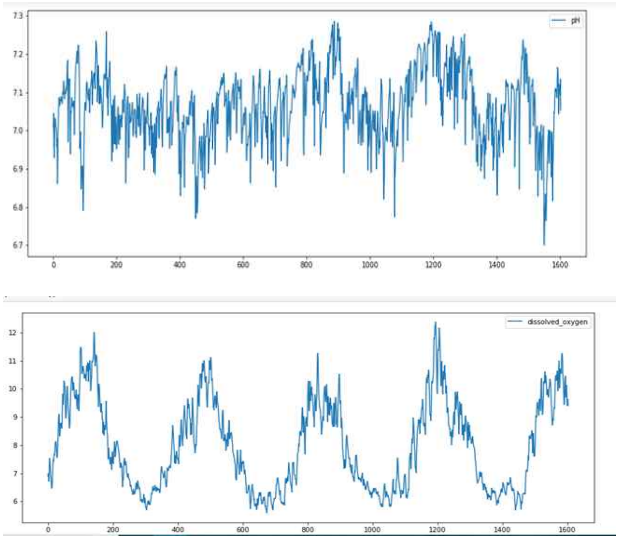

그림 10. pH 및 DO에 대한 측정 결과
Fig. 10. Measurement results for pH and DO

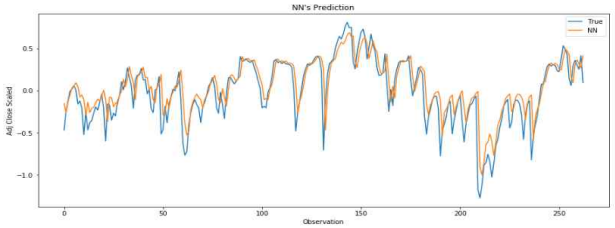

그림 11. 테스트 결과
Fig. 11. Train and Test Result

As a result of measuring by applying the WQ-NN Model, the data reliability that satisfies the quality control standard was obtained with less training data than the existing LSTM Model and ARIMA Model. Figure 11 below shows the Python execution result and final result developed as an algorithm for the WQ-NN Model.

Ⅳ. Conclusion

In this paper, a WQ-NN model was proposed for the quality control of water data, and as a result of applying this model, it has brought reliability in verifying data quality that meets quality control standards with less training data than existing LSTM Model and ARIMA Model Is showing. In verifying the accuracy of water quality data, it is not easy to verify whether the detected water quality data is normal or error because various factors in measuring water quality are complex and work, and the natural environment dynamically changes over time. Systematic analysis was not possible due to the lack of formal mathematical rules, and the integrity of the data was determined according to the situational judgment of experts. Therefore, as a solution to this

problem, middleware construction using a machine learning model based on big data is the best solution. In this study, a method for designing a real-time water quality data verification middleware S/W based on a multilayer neural network and developing it as a analyzing system was proposed.

Acknowledgment

This work was financially supported by the Research Year of Chungbuk National University in 2020.

References

[1] U.S. EPA Water Sentinel System Architecture Draft, Version 1.0 U.S EPA Water Security Division. J. Kang, “A Study on Analysis of Intelligent Video Surveillance Systems for Societal Security, 2005.

[2] NoSuk Park, Seonha Chae, Sukmin Yoon. A Study on the Statistical Predictability of Drinking Water Qualities for Contamination Warning System, Journal of Korean Society of Water and Wastewater Vol. 29, No. 4, 2015.

[3] NoSuk Park, SunHo Kim, Seonha Chae, Sukmin Yoon, A Study on the Turbidity Estimation Model Using Data Mining Techniques in the Water Supply System., Journal of Korean Society of Environment Engineering, Vol. 38, No. 2, 87~95, 2016.

[4] SukHoon Kim, Sung Kyoung. “A Water Environment Monitering System using the RISC Sensor Network Node”, The Korean Navigation Institute Journal Vol. 12 No. 2 pp.109-114, 2008.

[5] Daehyun Kwon, Seongjae Lee, Soosun Cho, “A Development of Water Sensor Data Generator”, 2009 Proceeding of the Korean Multimedia Society Fall Conference Vol. 12 No. 2, Korea Multimedia Society

[6] Androscinski. Distributed Operating system, The Logical Design, Wesley, 1991

[7] Coulouris, G.,Dollimore, J., Kindberg.T. Distributed Systems Concept and Design, 3nd edition, Addison-Wesley, 2014.

[8] Kim Bae-hyun, Kwon Moon-taek. A Study on Web Service Security, Information Security Vol.4 No.2, pp8, 2004.

[9] Joe Clabby. Web Service Gotchas, North American Operations Bloor Research, 2002.

[10] Bijoy Majumdar. Enhance UDDI to manage Web services. 11. 30, 2006.

[11] Francesco Salamone, Lorenzo Belussi, Cristian Currò. Application of IoT and Machine Learning techniques for the assessment of thermal comfort perception. Energy Procedia, Volume 148, pp. 798-805, 2018.

[12] Amirhossein Farahzadi, Pooyan Shams. Middleware technologies for cloud of things: a survey. Digital Communications and Networks, Volume 4, Issue 3, pp. 176-188, 2018

[13] Furqan Alam, Rashid Mehmood. Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT), Procedia Computer Science, Volume 98, pp. 437-442, 2016

박성훈(Sung-Hoon Park)

1982년 : 고려대학교 정경대학 통계학과 (경제학사)
1992년 : 인디애나 대학교 대학원 (공학석사)
1994년 : 고려대학교 대학원 (박사-분산컴퓨팅)
1982년～1989년: 두산정보통신 선임연구원
1994년～1996년: 두산정보통신 기술연구소 소장
1996년～2004년: 남서울대학교 컴퓨터공학과 교수
2004년～현 재: 충북대학교 컴퓨터공학과 교수
※관심분야 : 분산/모바일 컴퓨팅, 컴퓨터이론, 분산로봇틱스 등

유수창(Su-Chang Yoo)

2012년 : 충북대학교 전자정보대학 컴퓨터공학과 (공학학사)
2014년 : 충북대학교 전자정보대학원 컴퓨터공학과 (공학석사)
2016년 : 충북대학교 전자정보대학원 컴퓨터공학과 (박사수료)
2005년～2007년: 투원정보시스템 선임 연구원
2008년～2010년: 진주제일병원 전산팀 선임 연구원
2016년～현 재: 충북대학교 전자정보대학원 컴퓨터공학과 수료 후 등록
※관심분야 : 분산컴퓨팅, 병렬처리, 원격 리모팅, 데이터베이스 튜닝 등

방승범(Seung-Peom Pang)

1992년 : 단국대학교 공과대학 화학공학과 (공학학사)
2001년 : 광운대학교 대학원 컴퓨터공학과 (공학석사)
2019년 : 충북대학교 전자정보대학원 컴퓨터공학과 (박사수료)
2005년～2007년: 엠차지정보기술 연구소 연구 소장
2008년～2010년: ㈜GMC 기술연구소 소장
2016년～현 재: 엠팩엔지니어링 연구소장
※관심분야 : 분산컴퓨팅, 인공지능, 음성인식 등