

사물인터넷 환경에서 IoT 데이터 정합성 연구

최창원*

한신대학교 컴퓨터공학부 교수

A Study On IoT Data Consistency in IoT Environment

Changwon Choi*

Professor, Division of Computer Engineering, Hanshin University

요약 사물인터넷의 발달에 따라 IoT 기기에서 처리되는 데이터의 정확성도 중요시되고 있다. 사물인터넷에서 생산되는 데이터는 센서마다 다양한 포맷과 프로토콜을 사용하고 있기에 수집된 센서 데이터에 이상이 발생하면 정규화하고 통합하는 과정에서 데이터 오류로 인해 실패하거나 잘못된 데이터를 구성하게 된다. 사용자의 상황이나 IoT 기기의 이상 증상은 정확하게 판단되지 않기 때문에 사용자 서비스 장애가 발생하거나 실패하는 문제가 발생한다.

본 논문은 IoT 환경에서 발생하는 다양한 형태의 데이터가 IoT 기기의 특성을 반영하여 정상적인 범주 내에서 변화되는지를 수학적 함수로 산출하여 데이터의 정합성을 탐지하는 방법을 제안한다. IoT 데이터의 발생 특성을 파악하기 위해 '기울기 분석'을 활용한 방법과 '선형 회귀 분석'을 활용한 방법을 각각 제시하고 실험하였다. 기울기를 활용하는 방법은 '증가하는 속도'가 다음에 일어나는 현상에 영향을 미치는 IoT 데이터(센서 기기)에 적합하며, 선형 회귀를 활용하는 방법은 선형적으로 데이터가 움직일 때 '선형 회귀 함수로부터의 차이'가 다음에 일어나는 현상에 영향을 미치는 데이터(수도, 전기 계량기)에 적합하였다.

주제어 : IoT 데이터 정합성, 기울기 분석, 선형 회귀 분석, 스마트 관리 시스템

Abstract As the IoT technology is more developed, it is more important for the accuracy of IoT data. Since the IoT data supports a different formats and protocols, it is often happened that the IoT system is failed or the incorrect data is generated with the unreliable IoT devices(sensor, actuator). Because the abnormality of IoT device or the user situation is not detected correctly, this problem makes the user to be unsatisfied with the IoT system.

This study proposes the decision methodology of IoT data consistency whether the IoT data is generated in normal range or not by using the mathematical functions('gradient descent function' and 'linear regression function'). It may be concluded that the gradient function method is suitable for the IoT data which the 'increasing velocity' is related with the next generated pattern(eg. sensor devices), the linear regression function method is suitable for the IoT data which the 'the difference from linear regression function' is related with the next generated pattern in case the data has a linear pattern(eg. water meter, electric meter).

Key Words : IoT Data Consistency, Gradient Analysis, Linear Regression Analysis, Smart Management System

*교신저자 : 최창원(won@hs.ac.kr)

접수일 2022년 8월 24일

수정일 2022년 10월 14일

심사완료일 2022년 10월 18일

1. 서론

1.1 연구 배경 및 목적

최근 사물인터넷(IoT : Internet of Things) 서비스의 발전은 다양한 상황을 판단할 수 있는 데이터를 생산하고 스마트 장치를 이용하여 하나의 연결된 사회로 구성하고 있다. 사물인터넷은 다양한 산업 분야에서 센서와 인식 장치를 이용하여 자동화된 서비스로 편리함을 제공하는 핵심기술로 발전하고 있다. 주변의 장치와 연결된 상태에서 사용자를 위해 인식 장치 간의 자동 설정, 자율제어, 최적의 운영 상태를 유지하는 시스템 모니터링까지 복잡한 서비스 구조로 발전하고 있다. 사물인터넷에서 생산되는 데이터는 센서마다 다양한 포맷과 프로토콜을 사용하고 있기에 수집된 센서 데이터에 이상이 발생하면 정규화하고 통합하는 과정에서 데이터 오류로 인해 실패하거나 잘못된 데이터를 구성하게 된다. 센서의 고장은 이동 중인 사용자의 상황을 판단하는 데이터의 유실로 인해 지역 간의 서비스 연결성을 보장하지 못하는 문제와 그로 인한 맞춤형 개인화 서비스의 실패로 확장될 수 있다. 그리고 무인 자동차나 자율주행과 같은 스마트 서비스 시스템의 경우에는 이상 장애로 인해 장치의 제어가 되지 않아 인명 사고로 이어질 수 있다. 따라서 IoT 환경에서 수집되는 센서의 데이터에 대한 이상 유무를 실시간으로 판단하고 모니터링하는 것이 중요하다. 이러한 상태에서 이상 상태를 예측함으로써 장애나 사고의 예방의 중요성도 높아지고 있다[5].

본 논문은 IoT 환경에서 발생하는 다양한 형태의 데이터가 IoT 기기의 특성을 반영하여 정상적인 범주내에서 변화되는지를 수학적 함수로 산출하여 데이터의 정상성을 탐지하는 방법을 제안한다.

1.2 관련 연구

수학적 정의에 따르면 함수의 극대점(極大點, local maximum point)은 주위의 모든 점의 함수값 이상의 함수값을 갖는 점이다. 극댓값(極大값, local maximum (value))은 극대점이 갖는 함수값이다. 이는 IoT 기기들이 정상적으로 동작한다면 IoT 데이터가 일정한 범위(극댓값과 극솟값)내에서 변동되며 이를 통해 특정 함수로 변환시킬 수 있으며 이 함수를 통해 앞으로 발생하는 데이터의 예측이 가능할 수 있다[11].

사용량 기반의 보험(Usage-Based Insurance, UBI) 상품은 운전한 만큼 비용을 낸다는 접근으로, 속도와 위치, 기타 요소를 추적해 위험성을 평가하고 자동차 보험

료를 계산한다. 만약 IoT 기기가 제공하는 데이터가 틀리거나 파악할 수 없는 사태가 발생하는 경우이다. 그리고 고객과 접하는 IoT 기기가 생성한 데이터의 소유권이 어디에 있는지, 데이터에 문제가 생기면 누가 책임을 질 것인지도 명확하지 않다[1][13-15].

연구 [3]은 오류 판단 대상 시스템의 데이터 수집 단계에서 발생된 오류를 감지하고 빅데이터 기술을 활용하여 유사한 패턴을 갖는 설비(기기)의 고장 가능성을 예측하였다. 하지만 이 연구는 IoT 환경에서 사용되는 다양한 기기들의 특성을 개별적으로 고려하여 적용하기에는 활용성이 제한되는 문제점을 갖고 있다. 연구 [4]는 뉴럴 네트워크 기반 알고리즘을 이용하여 고장 탐지 기법을 제안하고 이를 통해 초기 오류 탐지를 하도록 하였지만 많은 경험 학습 데이터를 기반으로 하기에 초기에 적용하기에는 쉽지 않은 문제점을 갖고 있다. 연구[2]는 수전 설비의 전력 공급 변압기에서 측정된 전류 및 전압을 감지하여 사용자 설비의 전기적 고장에 의한 파급 장전 발생 여부를 판단한다. 이 연구는 정보 수신부, 사용자 정보를 전달하는 중계부 및 고장이 발생한 설비의 사용자 정보를 관리하는 서버를 포함하기에 일반적인 IOT 환경에 적용하기에는 초기 비용이 많이 발생하는 어려움을 갖고 있다. IoT 환경에서 이상 탐지를 위해 데이터 표준화 전처리와 시계열 기반의 SVM(Support Vector Machine) 모델을 적용하여 전 처리되지 않은 데이터와 전처리된 데이터의 이상 장애를 탐지하고 예측하였다[6].

2. 연구 방법 및 내용

2.1 기울기 분석을 활용한 데이터 정상성

IoT 데이터간의 '극값'은 매우 유용한 정보로 사용이 가능하다. 온도, 조도, 습도 등과 센서들은 일정 범위내에서 상승하다가 하락하게 되는 전환점으로 적용하고 이를 IoT 데이터에 활용한다면 극값의 발생이나 변화를 예



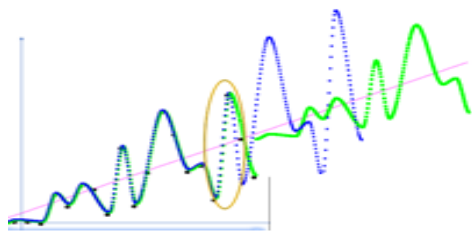
[Fig. 1] The Gradient Pattern

측할 수 있다[12]. 각 극값들을 연결하게 되면 기울기 (gradient)를 구할 수 있고 이는 기울기 변화에 맞맞은 함수를 구할 수 있게 된다[Fig 1].

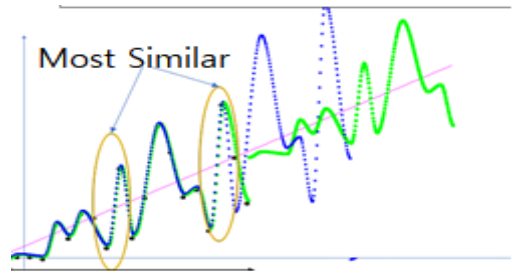
기울기를 활용하는 방법은 각 key 값에 따른 value 값들의 평균을 구하고 그 평균 중 극값이라고 생각되는 평균값만 산출하여 최근의 바로 전 구간의 기울기와 과거의 기울기를 통하여 유사한 구간을 찾아내고 다음 극값이 언제 일어날지를 추측하여 추가한 뒤 3차 함수의 미적분을 통하여 연결한 모형이다.

기울기를 활용하는 방법은 다음과 같다.

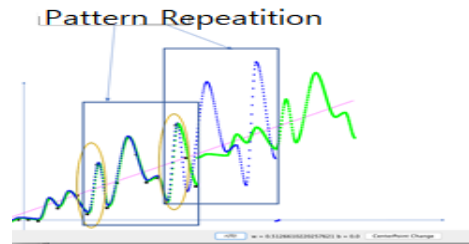
- 1) 데이터들을 key 값 기준으로 오름차순 정렬한다.
- 2) 데이터들의 key 값별로 군집화하여 value 값들을 모두 평균을 낸 (key, avgValue) 형태로 만든다.
- 3) 산출된 평균 좌표들을 차례로 스캔하며 avgValue 값이 증가하다가 감소하게 될 때 또는 감소하다 증가하게 될 때를 확인해 좌표들을 '극값'으로 정한다.
- 4) 모여진 극값 중 가장 최근에 발생한 극값의 특징을 선정한 뒤 그 특징 값을 기준(pivot)으로 삼는다.
 - 4-1) 가장 최근의 극값 2개를 선정하여 그 두 개의 극값을 좌표로 보고 평균 기울기를 구하고 그 값을 기준(pivot)으로 한다.[Fig. 2]
- 4-2) 앞에서부터 나머지 극값들을 2개씩 선정하여 그 사이의 기울기를 각각 구한다.
- 4-3) 기준(pivot)과 4-2)에서 구해진 기울기들을 비교하여 가장 유사한 구간의 위치를 찾아내어 기준(pivot)의 뒤에 연결한다.[Fig. 3]
- 5) 나머지 극값들 또한 앞에서 선정한 특징 값을 각각 구해 기준(pivot)의 특징과 가장 유사한 위치를 찾아내고 그 위치의 뒤를 앞으로 발생될 데이터로 간주한다.
- 6) 각 극값들을 2개씩 잡아 3차 함수의 미적분을 통해 연결하여 적절한 함수 f(x)를 산출한다.[Fig. 4]



[Fig. 2] Pivot Setting by Extreme Value



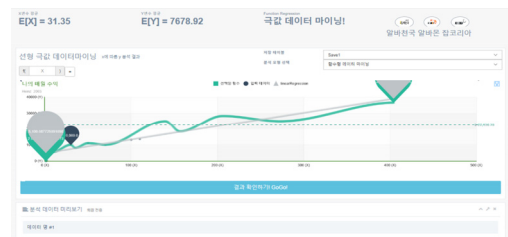
[Fig. 3] Search Similar Pattern



[Fig. 4] Function Calculation

2.2 선형 회귀 분석을 활용한 데이터 정합성

선형 회귀(linear regression)를 활용하는 방법은 데이터의 분포가 '이미 선형적으로 움직인다.'라는 것을 알고 있을 때 사용하는 분석 방식이며 전기, 수도 계량기 등의 다양한 IoT 기기들의 데이터 특성에 부합한다[Fig. 5]. 따라서 선형적인 데이터라는 것을 이용하여 선형적 분석 결과 함수와 얼마만큼 극값이 떨어져 있는지를 기준으로 하여 추측하고 과거의 패턴을 가져와 뒤에 연결하는 방식으로 분석한다.

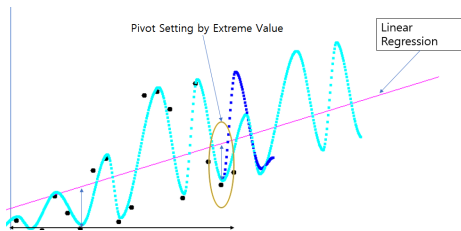


[Fig. 5] The Linear Regression Pattern

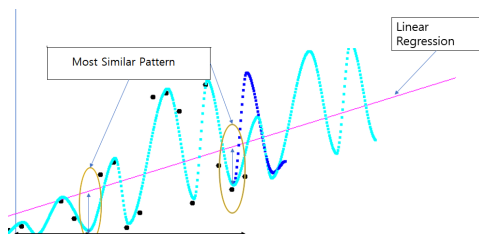
각 key값에 따른 value 값들의 평균을 구하고 그 평균값 중 극값이라고 생각되는 평균값만 추출하여 최근의 전 극값과 선형 회귀분석과의 차이와 과거의 극값과 선형회귀분석과의 차이를 비교하여 유사한 위치를 찾아내고 다음 극값이 언제 일어날지를 추측하여 추가한 뒤 3차 함수의 미적분을 통하여 연결한 모형이다.

선형 회귀를 활용하는 방법은 다음과 같다.

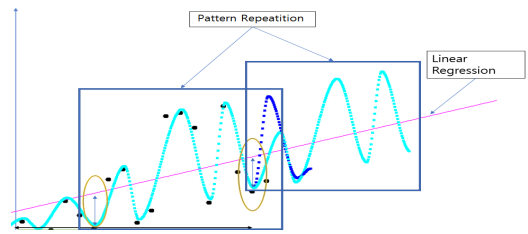
- 1) 데이터들을 key 값 기준으로 오름차순 정렬한다.
- 2) 데이터들의 key 값별로 군집화하여 value 값들을 모두 평균을 낸 (key, avgValue) 형태로 만든다.
- 3) 산출된 평균 좌표들을 차례로 스캔하며 avgValue 값이 증가하다가 감소하게 될 때 또는 감소하다 증가하게 될 때를 확인해 좌표들을 '극값'으로 정한다.
- 4) 모여진 극값 중 가장 최근에 발생한 극값의 특징을 선정한 뒤 그 특징 값을 기준(pivot)으로 삼는다.
 - 4-1) 가장 최근의 극값 1개를 선정하여 선형 회귀 분석 함수 $f(x)$ 로부터 얼마만큼 떨어져 있는지 계산하기 위해 $f(x_i) - y_i$ 값을 구하고 그 값을 기준(pivot)으로 삼는다.[Fig. 6]
 - 4-2) 맨 앞에서부터 다른 극값들이 선형 회귀분석 함수 $f(x)$ 로부터 얼마만큼 떨어져 있는지를 계산하기 위해 $f(x_j) - y_j$ 값을 구한다.
 - 4-3) 기준(pivot)과 4-2)에서 구해진 편차들을 비교하여 가장 유사한 극값 위치를 찾아내고 그 뒤를 기준(pivot)의 뒤에 연결한다. 이때 두개의 $f(x_i) - y_i$ 값을 닮음비로 간주하여 $f(x_j) - y_j : f(x_i) - y_i$ 비율로 크기를 조정한다.[Fig. 7]
- 5) 나머지 극값들 또한 앞에서 선정한 특징 값을 각각 구해 기준(pivot)의 특징과 가장 유사한 위치를 찾아내고 그 위치의 뒤를 앞으로 발생될 데이터로 간주한다.
- 6) 각 극값들을 2개씩 잡아 3차 함수의 미적분을 통해 연결하여 적절한 함수 $f(x)$ 를 산출한다.[Fig. 8]



[Fig. 6] Pivot Setting by Extreme Value



[Fig. 7] Search Similar Pattern



[Fig. 8] Function Calculation

앞에서 제안한 방법들 단계 6)의 3차 함수의 미적분을 통해 함수를 추출하는 이유는 다음과 같다.

- ① 직선은 기울기가 일정하여 미분이 불가능한 구간이 생긴다. 또한 한 번의 $f(x)$ 를 구하기 위한 시간 복잡도는 $(ax^1 + b) \times (\text{각 구간의 개수}) = O(n)$.
- ② 시간 복잡도가 극값이 n 개이며 그 값을 극값으로 갖는 다항 함수의 경우 극값이 n 개이고 극값으로 갖는 함수를 구해보면 $2n-1$ 차 함수가 생성되므로 함수 $f(x)$ 는

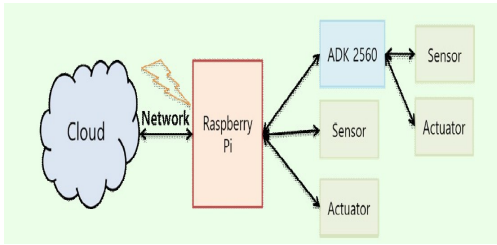
$$f(x) = ax^{2n-1} + bx^{2n-2} \dots + k$$
 이에 따라서 대략적인 연산 횟수 = $\frac{(2n-1)(2n)}{2}$
 이므로 시간 복잡도는 $O(n^2)$..
- ③ 3차 함수인 경우는 미분도 가능하며 시간 복잡도는 상수 시간 $(ax^3 + bx^2 + cx + d) \times (\text{각 구간 수}) = O(n)$.

따라서 미분이 가능해야 하며 IoT 데이터가 다양한 센서나 계량기 등에서 발생하는 '빅 데이터'라는 관점에서 처리 시간(시간 복잡도)을 고려하게 되면 3차 함수의 미적분을 통해 함수 $f(x)$ 를 추출하는 것이 가장 적합하다고 판단된다.

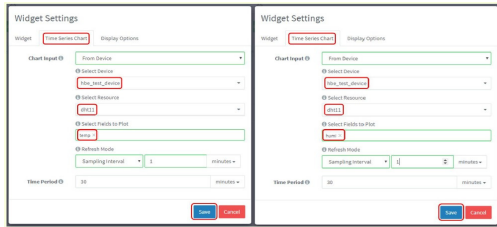
3. 실험 및 분석

본 논문에서 제안한 기울기 분석을 이용한 IoT 데이터 정합성 분석과 선형 회귀 분석을 이용한 IoT 데이터 정합성 분석을 위해 IoT 기기에서 발생된 데이터들을 수집하여 시각화하였다.[7][10]

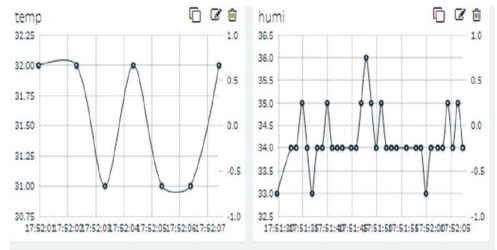
기울기 분석을 위해 [Fig. 9]와 같이 클라우드 환경 (Minute Cloud Platform)을 구축하고 라즈베리파이와 연동을 위해 JSON으로 API를 개발하였다[Fig. 10]. 데이터 생성을 위해 라즈베리파이와 온도/습도 센서를 연결하여 클라우드 상에서 시각화하도록 하였다[Fig. 11].



[Fig. 9] Experimental Environment



[Fig. 10] Cloud Interface



[Fig. 11] Generated Sensor Data

[Fig. 12]는 선형회귀 분석을 위해 클라우드 플랫폼을 이용하여 데이터들을 수집하고 극값을 추출한 후 도출된 함수를 시각화한 그림이다.



[Fig. 12] Visualization of $f(x)$

실험을 통해 기울기를 활용하는 방법은 ‘증가하는 속도’가 다음에 일어나는 현상에 영향을 미치는 IoT 데이터에 적합하며 선형 회귀를 활용하는 방법은 선형적으로 데이터가 움직일 때 ‘선형 회귀 함수로부터의 차이’(평균

에서의 떨어진 정도)가 다음에 일어나는 현상에 영향을 미치는 데이터에 적합한 것을 확인할 수 있었다.

4. 결론 및 향후 과제

사물인터넷은 다양한 산업 분야에서 센서와 인식 장치를 이용하여 자동화된 서비스로 편리함을 제공하는 핵심 기술로 발전하고 있지만 생산되는 데이터는 센서마다 다양한 포맷과 프로토콜을 사용하고 있기에 수집된 센서 데이터에 이상이 발생하면 정규화하고 통합하는 과정에서 데이터 오류로 인해 실패하거나 잘못된 데이터를 구성하게 된다. 따라서 IoT 환경에서 수집되는 센서의 데이터에 대한 이상 유무를 실시간으로 판단하고 모니터링 하는 것이 중요해졌으며 이상 상태를 예측함으로써 장애나 사고의 예방을 방지할 수 있다.

본 논문은 IoT 환경에서 발생하는 다양한 형태의 데이터가 IoT 기기의 특성을 반영하여 정상적인 범주내에서 변화되는지를 수학적 함수로 산출하여 데이터의 정합성을 탐지하는 방법을 제안하였다. 세부적으로 IoT 데이터의 발생 특성을 파악하기 위해 ‘기울기 분석’을 활용한 방법과 ‘선형 회귀 분석’을 활용한 방법을 각각 제시하고 실험하였다. 실험 결과 힘을 통해 기울기를 활용하는 방법은 ‘증가하는 속도’가 다음에 일어나는 현상에 영향을 미치는 IoT 데이터에 적합하며 선형 회귀를 활용하는 방법은 선형적으로 데이터가 움직일 때 ‘선형 회귀 함수로부터의 차이’(평균에서의 떨어진 정도)가 다음에 일어나는 현상에 영향을 미치는 데이터에 적합한 분석이라고 판단한다.

IoT 환경은 앞으로 산업 전반뿐만 아니라 특정 분야 등에도 활용될 것이며 이에 따른 다양한 센서, 액추에이터 등의 IoT 기기 사용이 늘어날 것이다. 아울러 다양하고 복잡한 IoT 기기들이 생성하는 데이터의 특성 또한 기존의 IoT 기기들과는 다른 형태를 나타낼 것이기에 향후 이에 대한 포괄적이고 정확한 데이터 특성의 분석 연구가 필요하다.

REFERENCES

[1] Fredric Paul , “IOT system error and the danger of bad IOT data,” IT world, Network World, 2019.5.
 [2] E. J. Joo et al., “Error Analysis for Temperature

BigData of Hydropower Collected by IoT sensors," pp.553-555, JKIPS, 2017.4

- [3] W. D. Cho et al., "Adaptive sensing and monitoring technologies for detecting big-data based on IoT multi-sensors," Journal of The Korea Institute of Information Scientists and Engineers, Vol.35, No.7, pp. 26-34, 2017.
- [4] K.T. Choi, "Neural networks optimization for multi-dimensional digital signal processing in IoT devices," Journal of Digital Contents Society, Vol.18, No.6, pp. 1165-1173, 2017.
- [5] D. H. Shin et al., "Trend and prospects internet of things," Journal of the Korea Society for Internet Information, Vol.14, No.2, pp.32-46, 2013.
- [6] Sung Il Na, "Design of Anomaly Detection System Based on Big Data in Internet of Things," Journal of Digital Contents Society, Vol.19, No.2, pp.377-383, 2018.2.
- [7] C. Choi et al., "A Simple Cost Analysis of Host ID-LOC Separating protocol using SDN Features," JKIOTS, Vol.2, No.4, pp.41-47, 2016.
- [8] K. Lee et al., "A Design on Learning Model using Triz on Project-based Learning in IOT," JKIOTS, Vol.5, No.3, pp.29-35, 2019.
- [9] S. Lee et al., "An Analysis of Software Development Process Based on Software Engineering in IOT Environment," JKIOTS, Vol.6. No.1, pp.25-31, 2020.
- [10] H. Lim et al., "A Design on Error Tracking System for Enhanced-Reliable IOT Service," JKIOTS, Vol.6. No.3, pp.15-20, 2020.
- [11] D. H. Shin et al., "Trend and prospects internet of things," Journal of the Korea Society for Internet Information, Vol.14, No.2, pp.32-46, 2013.
- [12] S. J. Lee, "Real time predictive analytic system design and implementation using big data-log," Journal of the Korea Institute of Information Security & Cryptology, Vol.25, No.6, pp.1399-1410, 2015.
- [13] Dong-Jin Shin et al., "Big Data-based Log Collection and Analysis in IoT Environments," Journal of Engineering and Applied Sciences, Vol.13, No.5, pp.1064-1072, 2018.
- [14] IoT Analytics Research, 2021.
- [15] www.lora-alliance.org

최 창 원(Chang-Won Choi)

[종신회원]



- 1990년 2월 : 고려대학교 전산과학과 졸업(학사)
- 1992년 2월 : 고려대학교 전산과학과 졸업(석사)
- 1995년 8월 : 고려대학교 전산과학과 졸업(박사)
- 1996년 ~ 현재 : 한신대학교 컴퓨터공학부 교수

<관심분야>

사물인터넷, 유무선 네트워크 및 보안, 시스템 분석