

제조 공정 빅데이터 분석을 위한 플랫폼 연구

구진희*

목원대학교 정보통신융합공학부

A Study on the Platform for Big Data Analysis of Manufacturing Process

Jin-Hee Ku*

Department of Information Communication Engineering, Mokwon University

요약 IoT, 클라우드 컴퓨팅, 빅데이터와 같은 주요 ICT 기술이 제조 분야에 적용되기 시작하면서 스마트 공장 구축이 본격화 되고 있다. 스마트 공장 구현의 핵심은 공장 내외부의 데이터 확보 및 분석력에 있다. 따라서 빅데이터 분석 플랫폼에 대한 필요성이 증가하고 있다. 본 연구의 목적은 제조 공정 빅데이터 분석을 위한 플랫폼을 구성하고, 분석을 위한 통합 메소드를 제안하는데 있다. 제안하는 플랫폼은 대량의 데이터 셋을 분산 처리하기 위해 분석도구 R과 하둡을 통합한 RHadoop 기반 구조로서 자동화 시스템의 단위 공정 및 공장 내에서 수집되는 빅데이터를 하둡 HBase에 직접 저장 및 분석이 가능하다. 또한 기존 RDB 기반 분석의 한계점을 보완하였다. 이러한 플랫폼은 스마트 공장을 위한 단위 공정 적합성을 고려하여 개발되어야 하며, 제조 공정에 스마트 공장을 도입하고자 하는 중소기업에 IoT 플랫폼 구축의 가이드가 될 수 있을 것으로 전망된다.

키워드 : 제조 공정, 빅데이터, 플랫폼, 사물인터넷, 스마트 공장

Abstract As major ICT technologies such as IoT, cloud computing, and Big Data are being applied to manufacturing, smart factories are beginning to be built. The key of smart factory implementation is the ability to acquire and analyze data of the factory. Therefore, the need for a big data analysis platform is increasing. The purpose of this study is to construct a platform for big data analysis of manufacturing process and propose integrated method for analysis. The proposed platform is a RHadoop-based structure that integrates analysis tool R and Hadoop to distribute a large amount of datasets. It can store and analyze big data collected in the unit process and factory in the automation system directly in HBase, and it has overcome the limitations of RDB - based analysis. Such a platform should be developed in consideration of the unit process suitability for smart factories, and it is expected to be a guide to building IoT platforms for SMEs that intend to introduce smart factories into the manufacturing process.

Key Words : Manufacturing Process, Big Data, Platform, Internet of Things, Smart Factory

1. 서론

인터넷이 보편화 된지 20년이 지난 현재, 인터넷은 IoT를 기반으로 우리가 살고 일하고 생산하는 방식을 완전히 바꿔놓았다. IoT가 제조업에 적용되기 시작하면서

개별적인 공장 설비들이 연결되고 기계의 지능화를 통한 최적의 공정을 위한 의사결정이 가능하도록 변화하는 이른바 스마트 공장 시대를 앞두고 있다. 스마트 공장은 제조 현장에서의 정보뿐만 아니라 기업의 비즈니스 관련하여 발생하는 모든 정보에 따라 최적의 의사결정을 내리

는 지능화된 공장을 의미한다[1]. 독일의 Industrie 4.0, 미국의 IIC(Industrial Internet Consortium)와 같은 국가적인 전략적 프로젝트에서도 스마트 공장 구축을 최우선 과제로 두고 있다[2]. 우리나라는 4차 산업혁명 시대에 제조업 혁신을 위해서 중소·중견기업의 스마트 공장 보급 확대 및 고도화 정책을 발표하였고, 2020년까지 1 만 개, 2025년까지 3 만개의 스마트 공장 구축을 발표하였다 [3].

이러한 국가적인 프로젝트들에서 그 핵심은 데이터 기반의 공장 운영체계를 구축하는 것이라고 할 수 있다. 따라서 성공적인 스마트 공장 구축을 위해서는 공장 내 외부의 데이터를 수집하고 분석하는 능력이 무엇보다 중요하다.

종래의 제조 자동화에서 이러한 스마트 공장으로 변화하기 위해서는 산업현장에서의 IoT와 인공지능 기술이 이뤄져야 하지만 현실적으로 스스로 의사결정을 하는 인공지능의 수준은 아니다. 현재는 대부분 기계들에 스마트 센서를 적용하고 센서로부터 수집된 데이터를 원격에서 분석하여 기계의 가동 상태, 기계의 파손 및 제품의 불량 발생 가능성 판단 및 예측, 원격 관리 등을 통한 장비의 효율 및 안정성 확대, 생산성 향상, 비용 절감 등에 초점을 두고 있다[2]. 또한 제조 공정에서 발생하는 데이터는 관계형 데이터베이스에 저장되어 통계적 공정 관리를 하는데 센서 데이터가 누적되면서 대용량 데이터셋을 기존의 관계형 데이터베이스에서 분석하기에는 어려움이 있다.

따라서 본 연구의 목적은 제조 공정에서 생성되는 빅 데이터 수집 및 분석 플랫폼을 제안하고 분석을 위한 통합 메소드를 구현하는데 있다.

본 논문의 구성은 다음과 같다. 2장에서는 자동화 시스템과 통계적 공정 관리 그리고 스마트 공장 구축을 위한 플랫폼에 대한 관련연구를 기술한다. 3장에서는 제조 공정 빅데이터 통합 분석 플랫폼 구조를 제시하고, 4장에서는 연구 결과를 기술한다. 마지막으로 5장에서는 결론과 향후 연구에 대하여 기술한다.

2. 관련연구

2.1 자동화 시스템

공장 자동화(Factory Automation)는 컴퓨터와 로봇과 같은 장비를 이용해 공장 전체의 무인화 및 생산 과정의

자동화를 만들려는 것으로 PLC(Programmable Logic Controller) 및 네트워크를 이용한 설비 및 공정 제어 핵심 역할을 한다[4,5]. PLC에서 데이터를 수집하여 HMI(Human Machine Interface)와 RDB(Relational Database)등에 전송하기 위해서는 OPC(OLE for Process Control)라는 소프트웨어 인터페이스 표준이 사용된다. 서버/클라이언트 구조로 구현되는 OPC는 제조 데이터를 필요로 하는 다양한 애플리케이션과 시스템에 데이터를 전달, 저장, 가공하는 역할을 한다. 다음 Fig. 1은 현재 일반적인 이산형 제조 공정에서 이뤄지는 데이터 수집과 분석 흐름을 나타낸 것으로 공장 설비의 다양한 센서 데이터가 PLC에 전송되고 HMI를 통해 RDB에 저장되는 시스템 구조를 나타내고 있다.

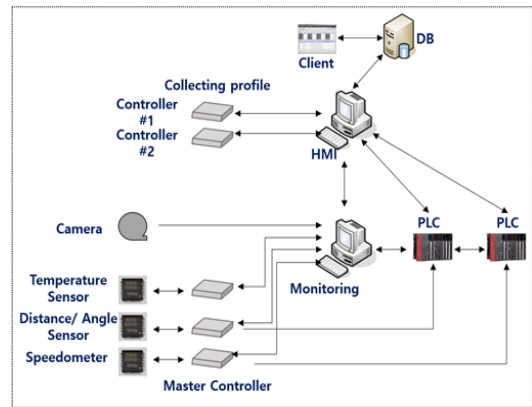


Fig. 1. Manufacturing process of automation system.

제조 산업 현장에서 이렇게 저장된 데이터는 통계적 자료와 분석기법으로 공정의 품질변동을 주는 원인과 공정의 능력 상태를 파악하여 품질 목표 달성을 위해 공정 개선이 이뤄지도록 관리하는 통계적 공정 관리 (Statistical Process Control)를 하고 있다. 공정 능력을 평가하기 위하여 일반적으로 사용되는 방법으로는 다음 식 (1)과 식 (2)의 C_p , C_{pk} 가 있다.

$$C_p = \frac{USL - LSL}{6\sigma} \quad (1)$$

$$C_{pk} = (1 - k)C_p \quad (2)$$

C_p 는 공정을 개선하기 위해서 요구되는 수준과 업무 결과에 대한 비교를 통해 공정능력을 측정하기 위한 방법으로 식 (1)과 같다. C_{pk} 는 C_p 가 중심으로부터 평균값

이 한쪽으로 치우쳐 있을 경우의 공정능력을 측정하는 방법으로 식 (2)와 같다[6,7]. 그러나 RDB를 분석 대상으로 하는 SPC의 문제점은 분석에 로드할 수 있는 ROW의 한계가 있고 ROW수가 증가할수록 효율은 현저히 저하된다는 점이다. 따라서 최근 공정이 복잡해지고 센서 데이터가 급증하면서 제조 공정 분석은 RDB를 대상으로 하는 SPC에서 대용량 데이터셋을 포함하는 빅데이터 분석으로 확대되어 가고 있다.

PLC를 기반으로 한 자동화 시스템은 컴퓨팅 기술과 센서들 그리고 산업용 기기들과 연동되면서 버려지던 공장 내외부의 데이터들을 축적하고 이를 기반으로 최적의 공정을 위한 더 나은 의사 결정을 가능하게 한다. 즉, 빅데이터 분석을 통해 공정과 공정을 담당하는 기계의 지능화가 가능한 것이다.

2.2 스마트 공장 구현을 위한 플랫폼

스마트 공장 구축의 목표는 가상 세계와 물리적 세계의 통합 시스템으로 사물들이 서로 소통하며 자동적, 지능적으로 제어되는 가상-물리 시스템(Cyber-Physical System)을 구현하는 것이다[8]. 스마트 공장 구축에 있어서 물리적 기기들을 확장성 있게 통합 제어하는 PLC의 구축은 매우 중요하다[5]. 최하위 레벨의 센서와 같은 펠드 장치가 아날로그 신호를 통해 PLC로 데이터를 보내고 HMI/SCADA 시스템은 원격 제어 작업을 수행한다. 상위 레벨인 MES(Manufacturing Execution Systems)는 사용자가 생산 일정 계획과 같은 복잡한 작업을 수행할 수 있도록 한다. 최상위 레벨의 ERP 시스템은 관리보고를 허용하고 다른 시스템과의 주문 상태와 같은 제조 데이터를 공유한다. 기업의 모든 제품과 개체들과의 관련된 정보 그리고 제조 시스템은 정형화된 정보로 RDB에 저장되게 되며, 가상 시스템의 여러 요소 및 ERP,

MES, PLM, CRM 등 응용프로그램들에서 다양하게 변형되어 사용되고 있다[9].

이와 같이 자동화 시스템은 지난 30 년 동안 Fig. 2의 왼쪽과 같이 5계층 자동화 아키텍처가 형성되었다. Fig. 2의 오른쪽은 스마트 공장 구축으로 실현되는 미래의 가상-물리 시스템을 나타낸 것이다.

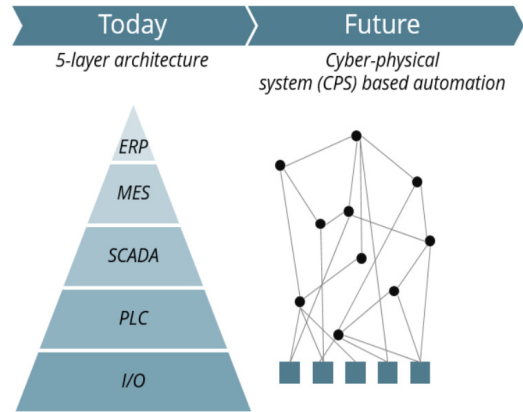


Fig. 2. Architecture of future automation system.

최근 PLC는 마이크로 컴퓨터를 심고 대용량, 고속화, 고성능 등 점차 발전된 형태로 진화하면서 제조 공정의 핵심으로 거듭나고 있다. 산업용 PC와 PLC가 자동화 시장에서 차지하는 비중은 아직도 60% 이상에 가깝다[2]. 2014년 산업 자동화 시스템 세계 시장 규모는 2014년 23조 3,200억원 이었으며, 2020년에는 45조 2,100억원이 될 것으로 전망하고 있다[10].

Table 1은 공장의 IT기술 활용정도 및 역량에 따라 스마트 공장 구축 레벨을 나타낸 것이다. Table 1에서 중간 단계는 센서와 PLC 등을 활용해 자동화 설비를 구축하고 실시간으로 제조 상황 파악이 가능한 공장이다. 즉, 스

Table 1. The level of smart factory by use of IT technology in factory.

Field / Level	Foundation stage	Intermediate stage	Advanced stage
Factory operation field (SW solution etc.)	Production performance management centered on process logistics	Quality analysis and real-time production management using facilities and production information	All stages for manufacturing by integration of planning-production-distribution logistics system
Equipment Automation (Sensor, Robot etc. HW)	Collect basic data using barcode and RFID	Automatic collection of real-time production information and equipment control using sensor, PLC, etc.	Multifunctional robot that can judge by oneself, the whole factory automatic operation with advanced CPS

마트 공장 구축을 위해서는 센서 및 필드 레벨에서 수집한 데이터가 통합되어 머신러닝 등 빅데이터를 통한 분석 작업이 필수적이다. 그러나 현재 제조 시스템을 둘러싼 정보의 흐름과 관리 상황은 개별적인 자동화 설비 및 이들의 단계적 도입으로 인해 통합되어 있지 않다.

3. 제조공정 빅데이터 분석 플랫폼

3.1 RHadoop기반의 통합 분석 플랫폼

기존의 자동화 시스템에서 공정이 복잡해지고 센서 데이터의 불량이 커지면서 기존의 RDB 기술만으로는 분석에 한계가 왔다. 센서 데이터와 같이 구조화 데이터일 지라도 데이터가 거의 실시간으로 유입되어 데이터양이 짧은 시간에 엄청나게 증가한다면 관계 데이터베이스는 데이터에 대한 쿼리 응답시간이 허용 범위를 넘기게 되기 때문에 처리가 곤란하다[11]. 또한 센서 데이터가 늘어나면서 분석 속도는 급격히 느려지는데 계속해서 RDB를 추가할 수는 없다. 따라서 대량의 데이터를 안정성 있게 저장할 수 있는 분산처리 기술 기반의 데이터 관리 기술들이 요구되고 있으며 기존의 제조 데이터베이스와도 연동이 가능한 플랫폼이 요구된다.

2006년 등장한 하둡은 현재까지 가장 널리 알려진 대용량 데이터 분산 처리 프레임워크이다. 하둡은 2개의 필수 요소를 갖고 있다. HDFS(Hadoop Distributed File System)과 MapReduce이다. 하둡은 기본적으로 분산형 데이터 기반으로 대량의 데이터 컬렉션을 서버 클러스터 내의 여러 개 노드들에 분산시키는 역할을 한다[12]. 하둡 시스템에서 작은 양의 데이터는 단일 노드 환경이, 많은 양의 데이터는 분산 노드 환경(N개의 데이터 노드로 구성 가능)이 좋은 성능을 발휘한다.

데이터가 기하학적으로 증가하였을 때 RDB에 저장한 데이터를 SQL로 처리하는 경우와 HBASE로 적재한 후 하둡의 하이브(Hive)를 이용한 맵리듀스 병렬처리로 연산하는 경우의 성능 비교에서 관계형 데이터베이스 시스템에서 SQL을 사용하는 경우 대부분의 연산이 하둡을 사용했을 경우보다 3~5배 이상 더 소요되었다[13]. 이는 HBASE와 같은 컬럼 기반 저장 방식에서는 동일한 연산 수행 시 디스크 입출력이 RDB보다 줄어들기 때문이다.

또한 기존 RDB와 하둡 위에 분석 솔루션을 통합한 경우에는 좀 더 좋은 성능 분석 결과를 보였다. 삼성SDS는 반도체 공정에서 수집되는 데이터가 불량 패턴 데이터와

얼마나 차이가 나는지 비교 분석하여 불량률을 분석하는데 걸리는 시간이 기존 방법과 비교해 평균적으로 240분에서 37분으로 6.5배 개선되었고, 같은 시간 당 불량 분석 처리 건수는 2건에서 13건으로 늘어났음을 제시하였다 [14]. Fig. 3은 R과 하둡의 통합 시스템 구조도를 보여주고 있다.

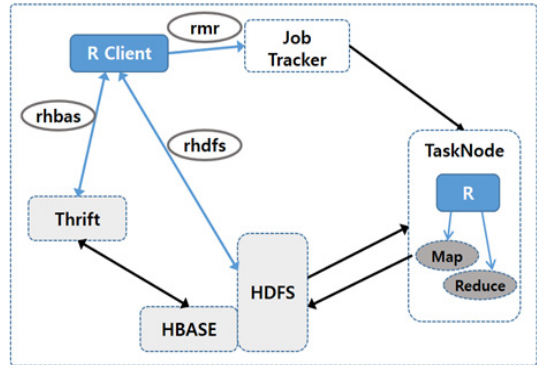


Fig. 3. Structure of RHadoop system.

Fig. 3의 시스템 구조도에서 R은 데이터 적재, 분석, 시각화 같은 데이터 분석 작업을 담당하고, 하둡은 데이터 병렬 저장소와 분산된 데이터에 대한 연산 작업을 담당한다.

4. 결과 및 논의

4.1 제조 공정 빅데이터 수집 및 분석 플랫폼 구조

본 연구에서는 하둡 클러스터의 능력을 결합하여 R의 처리 능력을 확대할 수 있는 방법으로 RHadoop기반 분석 플랫폼을 제안하였다. RHadoop기반 분석 플랫폼은 대용량 데이터를 저장한 곳에서 분석까지 할 수 있기 때문에 비용과 속도 면에서 효율적이다. 또한 PLC와의 통신을 위해 필요한 OPC 소프트웨어가 필요 없고, 스마트 센서가 클라우드로 직접 데이터를 전송하는 구조로 데이터 저장소를 기존의 RDB에서 하둡으로 변경하면 데이터 분석이 가능하다. 따라서 Fig. 2의 전통적인 자동화 시스템의 5개 레이어가 필요하지 않게 된다. 다음 Fig. 4는 하둡 위에 분석도구 R과 연동할 수 있는 플러그인을 추가한 RHadoop기반의 통합 분석 플랫폼 구조를 나타낸 것이다.

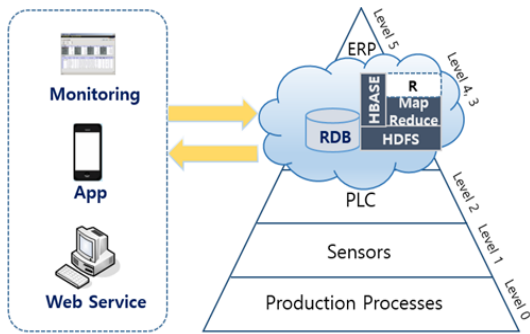


Fig. 4. RHadoop based integrated analysis platform structure.

위의 Fig. 4에서 제시된 플랫폼 구조는 기본적으로 데이터를 모든 중심에 놓는 것이 특징이다. 따라서 다양한 데이터베이스에 저장 가능하고 데이터를 사용하는 다양한 시스템과도 연동된다. 하둡은 R 이외에도 SAS 또는 SPSS와 같은 다른 통계 소프트웨어와도 통합 될 수 있다.

4.2 분석을 위한 데이터 통합 메소드

Table 2. Example of RHadoop based data analysis programming.

```
#Set the required Hadoop variable in RHadoop
Sys.setenv(HADOOP_HOME="usr/local/hadoop/")
Sys.setenv(HADOOP_CMD="usr/local/hadoop/bin/hadoop")
#Loading RHadoop library rmr2 and rhdfs
library(rhdfs)
library(rmr2)
#Initialize hdfs
hdfs.init()
#Defining large datasets
data <-
read.csv("usr/local/hadoop/sample_data/a.csv")
data.hdfs = to.dfs(data)
#Defining mapreduce job
mapreduce(
  input = data,
  map = function(k, v) {
    lapply(seq_along(v), function(r) {
      x <- runif(v[[r]])
      keyval(r, c(max(x), min(x)))
    })
  })
```

스마트 센서와 PLC를 통한 레이어 통합은 하둡에 R과 같은 분석도구를 엮어 통합 메소드를 구현하는 것을 가능하게 한다. RHadoop은 R과 HDFS를 연결하는

rhdfs, R과 HBASE를 연결하는 rhbase, R과 MapReduce를 연결하는 rmr 등 R환경에서 대량의 데이터 작업을 제공하는 3개의 R패키지로 구성되어 있다. RHadoop은 R과 Hadoop을 통합함으로써 데이터 이동 없이 R 코드를 작성하고 Hadoop에서 배치 처리할 수 있는 새로운 환경을 제공한다[15,16]. 다음 Table 2는 최적 상품가격 분석을 위해 전처리 과정을 거친 CSV파일을 RHadoop으로 분석하는 과정의 일부다. 맵리듀스 프로그램을 실행해보기 위해 맵 매개변수를 생성된 랜덤 변수의 max와 min을 계산하는 함수로 정의한 것이다.

5. 결론

독일, 미국, 일본 등 선진국을 중심으로 노동력 감소, 인건비 상승과 같은 요인으로부터 제품 경쟁력 강화를 목적으로 스마트 공장 구축이 급속히 증가하고 있는 추세이다. 스마트 공장은 산업 IoT 구현을 근간으로 하기 때문에 데이터마이닝, 머신러닝 등 빅데이터 분석이 필수적으로 요구된다. 본 연구에서는 제조 공정에서 발생하는 빅데이터 분석을 위한 플랫폼을 구성하고 분석을 위한 통합 메소드를 제안하였다. 연구를 위해 현재 자동화 시스템의 현황을 살펴보고 전통적 자동화 시스템 레이어를 통합하는 RHadoop기반 플랫폼을 구성하였다. 이러한 연구는 제조 공정 혁신을 위해 스마트 공장을 구축하고자 하는 중소 제조 기업에 기초적 방안이 될 것으로 전망한다. 스마트 공장을 성공적으로 구현하기 위해서는 무엇보다도 현재 제조 공정에서 생성되는 데이터를 최대한 활용하는 것이 중요할 것이다.

REFERENCES

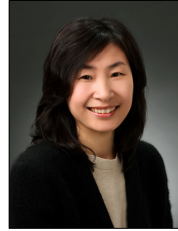
- [1] N. R. Lee. (2017). *Smart Factory 'just started' Technology secure is the key*. EPNC. <http://www.epnc.co.kr/news/articleView.html?idxno=72482>
- [2] J. H. Jang & J. H. Jeong. (2015). *The Smart Factory leading up to the Industrial Internet Revolution*. LG Business Insight.
- [3] Ministry of Trade, Industry and Energy. (2017). *Smart Manufacturing Innovation Vision 2025*.
- [4] BLOTTER. (2017). *Smart Factory is not Factory Automation*. BLOTTER. <http://www.blotter.net/archives/274222>

- [5] S. Y. Lee. (2015). *The role of IT for interacting with 'I' and the different 'you'(7) -Manufacturing as a Production Vs. Manufacturing as a Service-*. LG CNS. http://blog.lgcns.com/706#footnote_link_706_1
- [6] S. H. Park, Y. H. Park & M. J. Lee. (2005). *Statistical Process Control*. Seoul : MinYoungSa.
- [7] Wikipedia. (2017). *Process Capability Index*. Wikipedia. <https://ko.wikipedia.org>
- [8] Telecommunications Technology Association. (2017). *Information and communication terminology dictionary*. Telecommunications Technology Association. <http://terms.tta.or.kr>
- [9] IoT Analytics GmbH. (2017). *Will the industrial internet disrupt the smart factory of the future?*. IOT ANALYTICS. <https://iot-analytics.com/iot-market-segments-analysis/>
- [10] Technology and Information Promotion Agency for SMEs. (2015). *Small and Medium Enterprise Technology Roadmap 2017-2019*. Seoul : Technology and Information Promotion Agency for SMEs.
- [11] S. Makoto. (2013). *The Impact of Big Data*. Korea : HANBIT Media.
- [12] Apache Hadoop. (2014). *What Is Apache Hadoop?*. Apache Hadoop. <http://hadoop.apache.org>
- [13] Y. T. Cho, W. J. Lee, I. G. Lee, B. W. On & J. I. Choi. (2015). Analyzing smart grid energy data using Hadoop based big data system. *The Transactions of the Korean Institute of Electrical Engineers*, 64P(2), 85 - 91.
- [14] BLOTTER. (2017). *Samsung SDS applies 'Hadoop' to manufacturing process*. BLOTTER. <http://www.blotter.net/archives/158074>
- [15] J. H. Ku. (2017). A Study on Prediction Model of Equipment Failure Through Analysis of Big Data Based on RHadoop. *Wireless Personal Communications*, 80(1), 1-16.
DOI : 10.1007/s11277-017-4151-1
- [16] V. Prajapati. (2013). *Big data analytics with R and Hadoop*. Birmingham : Packt Publishing.

저 자 소 개

구 진 희(Jin-Hee Ku)

[정회원]



- 2001년 2월 : 충남대학교 컴퓨터과
학교육학과(교육학석사)
- 2010년 2월 : 충남대학교 공업(컴
퓨터)교육학과(교육학박사)
- 2010년 9월 ~ 현재 : 목원대학교
정보통신융합공학부 교수

<관심분야> : 컴퓨터과학 교육, 빅데이터, 데이터마이
닝, 머신러닝, 클라우드 컴퓨팅