

<https://doi.org/10.7236/JIIBC.2019.19.6.203>  
JIIBC 2019-6-29

## 빅데이터 기반 실시간 불량품 발생 원인 분석 및 설비 교체주기 예측

### Analysis of Defective Causes in Real Time and Prediction of Facility Replacement Cycle based on Big Data

황승연\*, 광경민\*, 신동진\*, 광광진\*\*, 노영주\*\*\*, 박경원\*\*\*\*, 박정민\*\*\*\*, 김정준\*\*\*\*\*

Seung-Yeon Hwang\*, Kyung-Min Kwak\*, Dong-Jin Shin\*, Kwang-Jin Kwak\*\*,  
Young-J Rho\*\*\*, Kyung-won Park\*\*\*\*, Jeong-Min Park\*\*\*\*, Jeong-Joon Kim\*\*\*\*\*

**요약** 최근 4차 산업혁명과 함께 세계 제조 강국들은 침체된 제조업의 부흥을 위해 국가적 전략을 추진하고 있다. 이러한 추세에 따라 문재인 정부도 '과학기술 발전이 선도하는 4차 산업혁명'이라는 전략을 제시하였다. 4차 산업혁명을 이끄는 핵심기술인 IoT, Cloud, Big data, Mobile, AI 등의 지능정보기술은 로봇, 3D 프린팅 등과 같은 신산업의 등장과 기존 주요 제조업의 스마트화를 촉진하고 있다. 스마트공장과 같은 기술이 발전함에 따라 IoT 기반의 센싱 기술이 발전하면서 이전에는 수집할 수 없었던 다양한 데이터를 측정할 수 있게 되었고, 각 공정에서 생성되는 데이터도 폭발적으로 증가했다. 따라서 본 논문에서는 데이터 생성기를 활용하여 스마트공장에서 발생할 수 있는 가상 데이터를 생성하고, 이를 활용하여 실시간으로 불량품의 발생 원인을 분석하고 설비의 교체주기를 예측하는 방법을 설명한다.

**Abstract** Along with the recent fourth industrial revolution, the world's manufacturing powerhouses are pushing for national strategies to revive the sluggish manufacturing industry. Moon Jae-in, the government is in accordance with the trend, called 'advancement of science and technology is leading the fourth round of the Industrial Revolution' strategy. Intelligent information technology such as IoT, Cloud, Big Data, Mobile, and AI, which are key technologies that lead the fourth industrial revolution, is promoting the emergence of new industries such as robots and 3D printing and the smarting of existing major manufacturing industries. Advances in technologies such as smart factories have enabled IoT-based sensing technology to measure various data that could not be collected before, and data generated by each process has also exploded. Thus, this paper uses data generators to generate virtual data that can occur in smart factories, and uses them to analyze the cause of the defect in real time and to predict the replacement cycle of the facility.

**Key Words** : Big Data, Data Analyze, Smart Factory, Smart Manufacturing, Virtual Data

\*준회원, 한국산업기술대학교 스마트팩토리융합학과 석사과정

\*\*준회원, 한국산업기술대학교 스마트팩토리융합학과 박사과정

\*\*\*정회원, 한국산업기술대학교 컴퓨터공학과 교수

\*\*\*\*정회원, 한국산업기술대학교 컴퓨터공학과 조교수

\*\*\*\*\*정회원, 한국산업기술대학교 컴퓨터공학과 조교수

접수일자: 2019년 10월 31일, 수정완료: 2019년 11월 30일

게재확정일자: 2019년 12월 6일

Received: 31 October, 2019 / Revised: 30 November, 2019 /

Accepted: 6 December, 2019

\*\*\*\*Corresponding Author: jkjm@kpu.ac.kr

Dept. of Computer Engineering, Korea Polytechnic University,

Korea.

## I. 서 론

최근 4차 산업혁명과 함께 제조업이 다시 떠오르면서, 세계 각국은 침체된 제조업을 부흥시키기 위한 국가적 전략을 추진하고 있다. 독일, 미국, 중국, 일본 등 제조 강국들은 생산 효율 증대와 친환경 고객 맞춤형 생산으로 제조업 경쟁력을 강화하고 있다. 이와 같은 추세에 따라, 문재인 정부도 2017년 7월 발표한 국정 5개년 계획에 '과학기술 발전이 선도하는 4차 산업혁명'이라는 전략을 제시하였다. 이처럼 국가별로 제조업 혁신을 위한 연구 및 기술개발을 확대하고 있으며, 각종 정책 및 프로그램을 운영하고 있다<sup>[1]</sup>.

4차 산업혁명을 이끄는 핵심기술인 IoT, Cloud, Big data, AI 등의 지능정보기술은 로봇, 3D 프린팅과 같은 신산업의 등장과 기존 주요 제조업의 스마트화를 촉진하고 있다<sup>[2]</sup>. 스마트공장은 기존의 제조업과 ICT(Information and Communication Technology) 기술을 융복합한 스마트제조 기술을 활용하여 새로운 가치 창출을 위한 종합 솔루션이라 할 수 있다<sup>[3, 4]</sup>. 스마트공장과 같은 기술이 발전함에 따라 IoT 기반의 센싱 기술이 발전하면서 이전에 수집하지 못했던 다양한 데이터를 측정할 수 있게 되었고, 각 공정에서 생성되는 데이터도 폭발적으로 증가했다<sup>[5]</sup>. 또한, Hadoop과 MapReduce와 같은 빅데이터 관련 오픈소스 기술의 발전으로 적은 비용으로 대량의 데이터를 저장하고 빠른 속도로 분석할 수 있게 되었다<sup>[6, 7]</sup>. 따라서 본 논문에서는 데이터 생성기를 활용하여 스마트공장에서 발생할 수 있는 가상 데이터를 생성하고, 이를 활용하여 실시간으로 불량품의 발생 원인을 분석하고 설비의 교체주기를 예측하는 방법을 설명한다.

본 논문은 다음과 같이 구성된다. 2장에서는 불량품 발생 원인 및 설비 교체주기를 예측하기 위한 빅데이터 기반의 분석 및 예측 관련 기술을 설명하고, 3장에서는 본 연구에서 생성한 가상 데이터의 생성 조건과 데이터 분석 시나리오를 설명한다. 4장에서는 가상 데이터를 활용하여 데이터 분석을 수행하고 그 결과를 설명한다. 마지막으로 5장에서 결론을 맺는다.

## II. 관련 연구

기존 제조업의 스마트화에 따라 스마트공장에서 사용하는 센서의 개수와 데이터량이 증가하였고 이러한 데이터를 분석하면 기존에는 발견하지 못했던 새로운 가치를

창출할 수 있다. 따라서 본 장에서는 여러 분석 기법 중, 본 연구에서 사용되는 시계열 분석, 군집 분석, 분류 분석, 딥러닝에 대해 알아본다.

### 1. 시계열 분석

시계열 분석이란 시간 변수의 흐름에 따른 종속변수의 움직임을 이해하고 예측하는 것을 목표로 하는 분석방법이다. 본 연구에서는 시계열 분석을 이용하여 공정에서 사용되는 소모성 부품의 교체주기를 예측한다. 본 연구에서는 시계열 데이터 분석 기법으로 RNN(Recurrent Neural Network)의 한 종류인 LSTM(Long Short Term Memory)을 사용하였다. LSTM은 장기 의존성 문제를 해결하기 위해 설계되었고, 전체 체인을 관통하는 셀 스테이트(Cell state)를 통해 이전 학습결과를 큰 변함없이 다음 단계로 전달한다. LSTM은 첫 번째 단계로 시그모이드(Sigmoid) 함수를 이용하여 삭제할 정보를 결정한다. 두 번째 단계는 어떤 정보를 셀 스테이트에 저장할지 결정하는 단계로 시그모이드 함수와 tanh 함수를 사용한다. 세 번째 단계에서는 셀 스테이트를 업데이트하고 어떤 값을 출력할지 결정한다<sup>[8]</sup>.

### 2. 군집 분석

군집화(Clustering)는 비지도 학습의 대표적인 방법으로, 유사한 요소를 동일한 그룹으로 나누는 것을 의미한다. 이렇게 요소들을 그룹으로 나누는 과정을 통해 유사한 요소들은 서로 가깝게, 서로 다른 군집 사이의 요소 간에는 멀게 하는 것이 군집화의 목표이다<sup>[9]</sup>. 본 연구에서는 이러한 군집화를 이용하여 최종 불량률을 최소화하기 위한 최적의 조건을 찾는 군집 분석을 수행한다. 본 연구에서 사용한 군집 분석 기법으로 DBSCAN(Density-based spatial clustering of applications with noise)을 사용하였다. DBSCAN은 데이터들 사이의 밀도 차이에 기반을 두고 수행되는 알고리즘으로 군집 개수가 초기에 주어지지 않아도 효율적으로 군집화를 수행하고, 같은 데이터 집합에 대해 여러 번 수행 하더라도 항상 같은 결과를 도출하는 알고리즘이다<sup>[10]</sup>.

### 3. 회귀 분석

회귀 분석은 종속변수와 독립변수 간의 상관관계를 검증하여 독립변수가 종속변수에 어떠한 영향을 미치는지 파악하거나, 독립변수의 변화에 따라 종속변수의 변화를 예측하기 위해 사용되는 분석 기법이다<sup>[11]</sup>. 본 연구에서

는 OLS(Ordinary Least Squares) 회귀 분석을 통해 각 공정에서 생성되는 데이터를 이용하여 불량률을 예측한다.

### III. 본 론

본 장에서는 데이터 생성기(Data Generator)를 이용하여 스마트공장에서 생성될 수 있는 데이터를 가상으로 생성하는 방법 및 조건에 대해 알아보고, 생성한 가상 데이터를 활용한 데이터 분석 시나리오를 설명한다.

#### 1. 가상 데이터 생성

본 연구에서는 하나의 완제품을 만들기 위해 A, B, C 공정으로 구성된 시나리오를 활용하여 가상 데이터를 생성한다. 본 절에서는 데이터 생성기를 이용하여 생성되는 각 공정의 가상 데이터 생성 방법 및 조건에 대해 알아본다.

##### 가. A 공정

A 공정에서 생성되는 데이터는 부품 ID, 시간, 가열온도(Heating temperature), 주입속도(Injection Speed), 제품완성도, 공기 흡입 컵(Air Suction Cup)의 크기, 불량 여부, 불량 개수, 생산량, 불량률이다. 제품의 완성도는 가열온도와 주입속도에 따라 결정되며, 부품의 불량 여부에 영향을 미친다. 예를 들어, 제품의 완성도가 50%일 경우, 생성된 부품이 양품일 확률이 50%이다. 가열온도와 주입속도에 따른 제품완성도의 데이터 생성 조건은 표 1, 표 2와 같다. 각 부품의 불량 여부는 환경 온도와 환경 습도의 제품완성도 중 더 낮은 값의 영향을 받도록 설정하였다. 예를 들어, 부품 A 생성 당시의 환경 온도가 21이고 습도가 45인 경우, 해당 환경 습도의 제품완성도는 100%이지만 환경 온도의 제품완성도는 50%이므로 해당 부품은 50%의 확률로 불량 여부가 결정된다.

표 1. 데이터 생성 조건 - 가열온도

Table 1. Data Generation Conditions - Heating Temperature

가열온도	제품완성도
1000	50%
1001~1050	70%
1100~1300	100%
1301~1350	70%
1400	50%

표 2. 데이터 생성 조건 - 주입속도

Table 2. Data Generation Conditions - Injection Speed

주입속도	제품완성도
60~70	50%
71~80	70%
81~100	100%
101~110	70%
111~120	50%

초기 공기 흡입 컵의 크기는 10mm로 설정하였고, 5mm가 되면 교체하도록 설정하였으며 제품완성도에 따라 감소하는 크기가 달라지는데, 그 조건은 표 3과 같다. 제품완성도가 100%일 경우, 0.0001mm, 70%일 경우 0.0003mm, 50%일 경우 0.0005mm씩 감소하도록 설정하였다.

표 3. 제품완성도에 따른 공기 흡입 컵의 크기 감소량

Table 3. Amount of reduction in size of air intake cup according to product completeness

제품완성도	공기 흡입 컵의 크기 감소량
100%	0.0001mm씩 감소
70%	0.0003mm씩 감소
50%	0.0005mm씩 감소

##### 나. B 공정

B 공정에서 생성되는 데이터는 부품 ID, 시간, 압력 수치(Holing pressure), 냉각 온도(Cooling temperature), 제품완성도, 불량 여부, 불량 개수, 생산량, 불량률이다. B 공정의 제품완성도는 압력 수치에 따라 결정되며, 부품의 불량 여부에 영향을 미친다. 압력 수치에 따른 제품완성도의 데이터 생성 조건은 표 4와 같고, 냉각 온도에 따른 제품완성도의 데이터 생성 조건은 표 5와 같다. 각 부품의 불량 여부는 압력 수치와 냉각 온도의 제품완성도 중 더 낮은 값의 영향을 받도록 설정하였다.

표 4. 데이터 생성 조건 - 압력 수치

Table 4. Data Generation Conditions - Pressure

압력 수치	제품완성도
400~500	50%
501~600	70%
601~800	100%
801~900	70%
901~1000	50%

표 5. 데이터 생성 조건 - 냉각 온도

Table 5. Data Generation Conditions - Cooling Temperature

냉각 온도	제품완성도
80	50%
81~84	70%
85~95	100%
96~99	70%
100	50%

#### 다. C 공정

C 공정에서 생성되는 데이터는 부품 ID, 시간, 경화 온도, 경화 습도, 경화 시간, 제품완성도, 불량 여부, 불량 개수, 생산량, 불량률이다. 제품의 완성도는 경화 온도와 경화 습도에 따라 결정되며, 부품의 불량 여부에 영향을 미친다. 경화 온도와 경화 습도에 따른 제품완성도의 데이터 생성 조건은 표 6, 표 7과 같다. 각 부품의 불량 여부는 경화 온도와 경화 습도의 제품완성도 중 더 낮은 값의 영향을 받도록 설정하였다.

표 6. 데이터 생성 조건 - 경화 온도

Table 6. Data Generation Conditions - hardening Temperature

경화 온도	제품완성도
21~22	50%
22~23	70%
24~27	100%
28~29	70%
29~30	50%

표 7. 데이터 생성 조건 - 경화 습도

Table 7. Data Generation Conditions - hardening Humidity

경화 습도	제품완성도
25~34	50%
35~44	70%
45~55	100%
56~70	70%
71~80	50%

#### 라. Result 공정

Result 공정은 A, B, C 공정에서 생성되는 부품들을 조립하여 완제품을 생성하는 공정으로, 생성되는 데이터는 날짜, A 부품의 불량 여부, B 부품의 불량 여부, C 부품의 불량 여부, 완제품 불량 여부, 불량 개수, 생산량, 불

량률이다. A, B, C 부품의 불량 여부는 각 공정의 불량 여부 데이터를 참고하고, 완제품 불량 여부는 각 공정의 부품 불량 여부를 통해 결정한다. 완제품의 불량 여부를 판단하는 조건은 표 8과 같다.

표 8. 완제품 불량 여부 판단 조건

Table 8. Conditions for determining product defect

공정별 부품 불량 여부	완제품 양품 확률
3개 양품(1, 1, 1)	100%
2개 양품(1, 1, 0)	70%
1개 양품(1, 0, 0)	50%

## 2. 데이터 분석 시나리오

본 절에서는 데이터 생성기에 의해 생성된 가상 데이터를 활용한 데이터 분석 시나리오에 대해 알아본다.

### 가. 시계열 분석

A 공정의 시간의 흐름에 따른 공기 흡입 컵의 크기 변화를 분석하여 공기 흡입 컵의 교체주기를 예측한다.

### 나. 군집 분석

A 공정에서는 가열온도, 주입속도, 제품완성도를 이용하고 B 공정에서는 냉각 온도, 압력 수치, 제품완성도를 이용하며 마지막으로 C 공정에서는 경화 온도, 경화 습도, 제품완성도를 이용하여 불량률을 최소화하는 최적의 군집을 찾기 위한 군집 분석을 수행한다.

### 다. 회귀 분석

A 공정에서는 불량률을 종속변수, 가열온도, 주입속도, 제품완성도를 독립변수로 설정하고 불량률을 예측한다.

B 공정에서는 불량률을 종속변수, 냉각 온도, 압력 수치, 제품완성도를 독립변수로 설정하고 불량률을 예측한다.

C 공정에서는 불량률을 종속변수, 경화 온도, 경화 습도, 제품완성도를 독립변수로 설정하고 불량률을 예측한다.

## IV. 분석 및 결과

### 가. 시계열 분석을 이용한 교체주기 예측

시계열 분석을 이용한 교체주기 예측을 하기 위해 다변수, 다층 레이어를 이용한 LSTM을 이용하였다. 사용된 변수는 A 공정의 가열온도, 주입속도, 공기 흡입 컵,

제품완성도 4가지이고, 이를 이용하여 시간의 흐름에 따른 공기 흡입 컵의 크기를 예측한다. 손실함수는 Mean-squared-error를 사용하고, Optimizer는 RMSProp을 사용하였다.

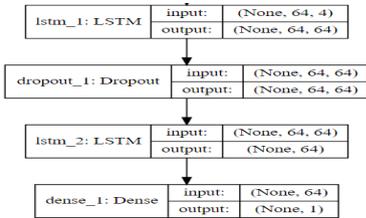


그림 1. LSTM 전체 구성도  
 Fig. 1. LSTM Overall structure

그림 1과 같이 구성된 레이어는 윈도우 사이즈를 64 개로 설정하고 윈도우 사이즈 이후의 수치를 예측하는 모델의 구성도이다. 과적합을 방지하기 위해 2개의 층 사이에 Dropout Layer를 추가하였다.

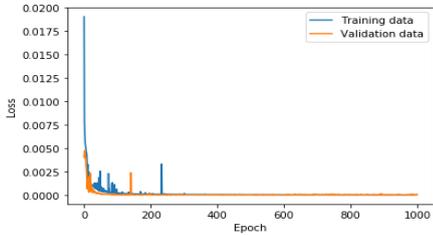


그림 2. 전체 학습 손실 값 그래프  
 Fig. 2. Overall training Loss graph

그림 2를 통해 전체 학습 Loss table을 확인할 수 있고 Train loss와 validation loss의 값이 같이 감소하는

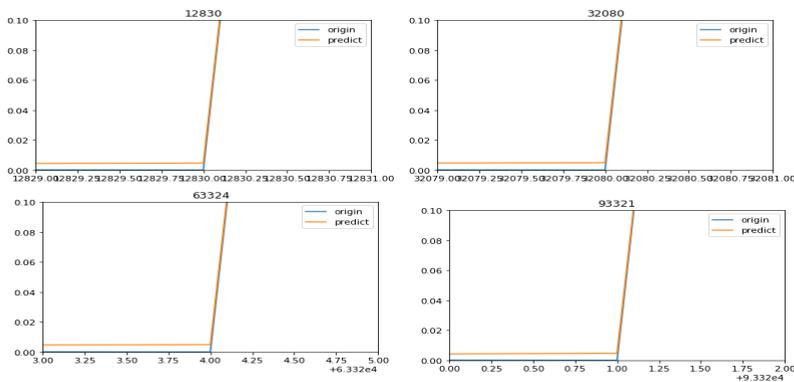


그림 5. 교체주기 부분을 확대한 샘플  
 Fig. 5. Exchanging periods samples

것을 통해 과적합이 되지 않은 것을 알 수 있다.

```

    ## evaluation loss and metrics ##
    299981/299981 [=====]
    Loss      : 7.258950320451952e-05
    
```

그림 3. 검증 셋을 통해 예측된 결과의 손실 값  
 Fig. 3. The value of the loss of the predicted results from the validation data.

그림 3을 통해 검증 데이터를 이용하여 예측된 결과의 손실이 0.00007로 아주 낮은 것을 확인할 수 있다.

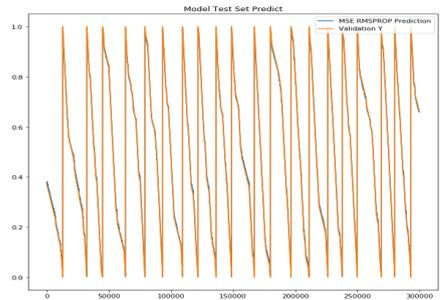


그림 4. 검증 데이터로 예측된 결과값  
 Fig. 4. Prediction value using validation data set

그림 4의 주황색으로 그려진 그래프는 훈련 데이터를 이용하여 학습한 모델이며 훈련에 사용되지 않은 검증 데이터를 예측한 그래프이고, 파란색으로 그려진 그래프가 실제 검증 데이터 값이다. 전체적인 모습은 일치해 보이며 그림 3을 통해 검증 데이터를 이용하여 예측된 결과의 손실이 0.00007로 손실 또한 낮은 편인 것을 확인할 수 있다.

그림 5는 공기 흡입 컵 교체주기를 4개를 샘플링 하여 확대한 그림이다. 이 그래프 또한 주황색이 예측한 데이터이고 파란색이 실제 데이터이다. 예측된 교체주기(상승 구간)와 실제 교체주기가 일치하여 학습된 모델로 교체 주기 예측이 가능하다.

나. 각 공정 데이터를 이용한 군집 분석

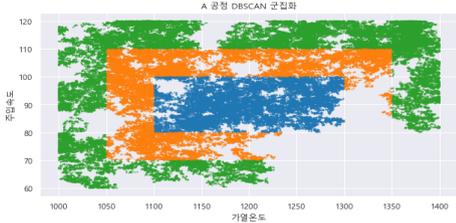


그림 6. DBSCAN을 이용한 밀도 기반의 A 공정 군집 분석  
Fig. 6. Cluster Analysis of Density-Based A Process Using DBSCAN

그림 6은 A 공정을 DBSCAN 알고리즘을 이용하여 밀도 기반의 군집 분석 결과이다.  $\epsilon$ 은 0.2, min samples는 1000을 가지는 DBSCAN 모델로, 여기서 epsilon은 거리를 의미하고, min samples는 epsilon 거리 안에서의 min samples 개수만큼을 제외한 나머지는 노이즈로 인식하는 것을 의미한다. 데이터는 군집화하기 이전에 0~1의 범위로 데이터를 스케일링한다. 군집화를 위해 가열온도, 주입 속도, 제품완성도 3가지의 변수를 사용한다. 3개의 군집이 나타나는데, 각 군집은 제품 완성도 별로 군집화되어있으며, 각 군집은 제품완성도에 따른 불량 확률을 보인다. 이를 이용하여 각각의 군집의 평균 가열온도와 주입 속도를 계산해 최적의 제품 생산 환경 분석이 가능하다. 파란색 군집은 100% 양품이 나오는 군집이고, 두 번째로 위치한 주황색 군집은 90% 양품을 갖는 확률이며, 세 번째로 위치한 초록색 군집은 50%의 양품을 갖는 군집이다. 각각의 군집은 그림 7과 같이 평균, 표준편차, 분산을 갖는다.

Blue Cluster Mean (Temperature, Speed):	1199.22, 90.21
Blue Cluster Var (Temperature, Speed):	2769.21, 28.74
Blue Cluster Std (Temperature, Speed):	52.62, 5.36
Orange Cluster Mean (Temperature, Speed):	1160.15, 96.64
Orange Cluster Var (Temperature, Speed):	8368.95, 157.22
Orange Cluster Std (Temperature, Speed):	91.48, 12.54
Green Cluster Mean (Temperature, Speed):	1196.30, 102.06
Green Cluster Var (Temperature, Speed):	18524.95, 329.02
Green Cluster Std (Temperature, Speed):	136.11, 18.14

그림 7. A 공정 각 군집의 평균, 표준편차, 분산  
Fig. 7. A Process Average, Standard Deviation, and Variance for each cluster

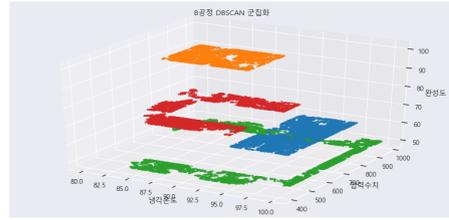


그림 8. DBSCAN을 이용한 밀도 기반의 B 공정 군집 분석  
Fig. 8. Cluster Analysis of Density-Based B Process Using DBSCAN

그림 8은 위와 같은 방법으로 B 공정을 군집 분석한 결과이다. C 공정과 차이점은 C 공정은 제품완성도가 3단계로 나누어지지만, B 공정은 4단계로 나누어진다. 따라서 2D 그래프를 사용할 경우 혼동이 생길 수 있으므로 3D 그래프로 대체하였다. 군집화를 위해 냉각 온도, 압력 수치, 제품완성도 3가지의 변수를 사용하고 각각의 군집은 그림 9와 같은 평균, 표준편차, 분산을 갖는다.

Blue Cluster Mean (Temperature, Speed):	96.26, 698.67
Blue Cluster Var (Temperature, Speed):	2.63, 11539.05
Blue Cluster Std (Temperature, Speed):	1.62, 107.42
Orange Cluster Mean (Temperature, Speed):	98.07, 696.84
Orange Cluster Var (Temperature, Speed):	5.54, 3983.99
Orange Cluster Std (Temperature, Speed):	2.35, 63.12
Green Cluster Mean (Temperature, Speed):	93.29, 689.90
Green Cluster Var (Temperature, Speed):	40.50, 43590.21
Green Cluster Std (Temperature, Speed):	6.36, 208.79
RED Cluster Mean (Temperature, Speed):	87.06, 675.23
RED Cluster Var (Temperature, Speed):	9.10, 17522.16
RED Cluster Std (Temperature, Speed):	3.02, 132.37

그림 9. B 공정 각 군집의 평균, 표준편차, 분산  
Fig. 9. B Process Average, Standard Deviation, and Variance for each cluster

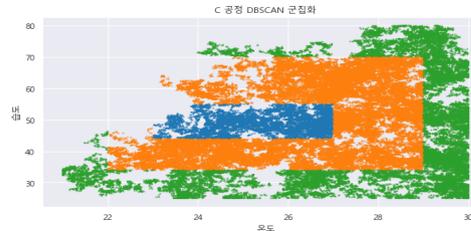


그림 10. DBSCAN을 이용한 밀도 기반의 C 공정 군집 분석  
Fig. 10. Cluster Analysis of Density-Based C Process Using DBSCAN

그림 10은 A, B 공정과 같이 epsilon, min samples 값을 주어 나타난 군집이다. 군집화를 위해 온도, 습도, 완성도 3가지의 변수를 사용하고 각각의 군집은 그림 11과 같은 평균, 표준편차, 분산을 갖는다.

Blue Cluster Mean (Temperature, Humidity): 25.50, 48.81  
 Blue Cluster Var (Temperature, Humidity): 1.14, 8.06  
 Blue Cluster Std (Temperature, Humidity): 1.07, 2.84  
 Orange Cluster Mean (Temperature, Humidity): 26.46, 50.49  
 Orange Cluster Var (Temperature, Humidity): 3.04, 126.52  
 Orange Cluster Std (Temperature, Humidity): 1.74, 11.25  
 Green Cluster Mean (Temperature, Humidity): 27.81, 45.64  
 Green Cluster Var (Temperature, Humidity): 5.78, 314.42  
 Green Cluster Std (Temperature, Humidity): 2.40, 17.73

그림 11. C 공정 각 군집의 평균, 표준편차, 분산  
 Fig. 11. C Process Average, Standard Deviation, and Variance for each cluster

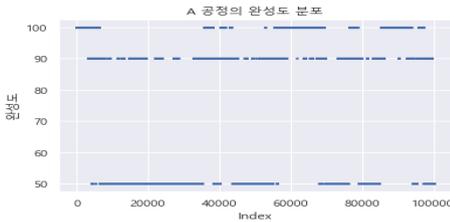


그림 12. A 공정의 제품완성도 분포  
 Fig. 12. Product Completion Table for Process A

그림 12는 A 공정의 제품완성도 분포로 50, 90, 100의 3가지 양상이 나타난다.



그림 13. B 공정의 제품완성도 분포  
 Fig. 13. Product Completion Table for Process B

그림 13은 B 공정의 완성도 분포로 50, 65, 70, 100의 4가지 양상이 나타난다.

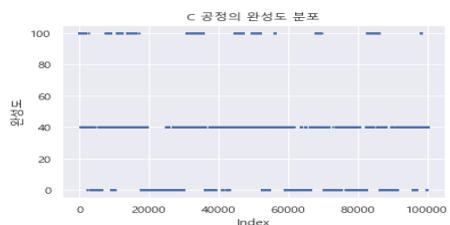


그림 14. C 공정의 제품완성도 분포  
 Fig. 14. Product Completion Table for Process C

그림 14는 A 공정의 제품완성도 분포로 0, 40, 100의 3가지 양상이 나타난다.

다. 각 공정 데이터를 이용한 회귀 분석

각 A, B, C 공정에서 생성되는 데이터를 활용하여 회귀 분석을 진행하였다.

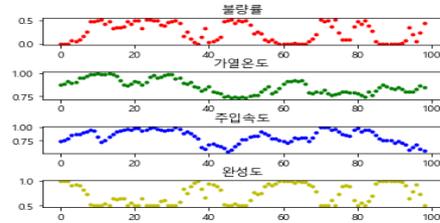


그림 15. A 공정 데이터의 window size별 평균값  
 Fig. 15. Average value by window size of process A data

그림 15는 총 10만 개의 데이터 중 불량률, 온도, 속도, 제품완성도를 window size(본 연구에서는 1000개로 지정하였다)마다 평균을 내어 그래프로 나타낸 것이다. 이중 독립변수는 가열온도, 주입속도, 제품완성도이며, 종속변수는 불량률에 해당한다. 불량률은 window size 크기의 데이터에서 불량 여부를 카운트하여, window size로 나눈 평균값이다. 가열온도, 주입속도, 제품완성도 또한 window size 만큼씩 더한 후 window size로 나눈 평균값이다. 위 그래프의 종속변수와 독립변수를 이용하여, OLS(Ordinary Least Squares) 회귀 분석을 진행한다.

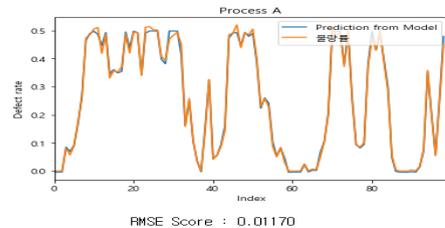


그림 16. A공정의 불량 예측 그래프와 RMSE 스코어  
 Fig. 16. Predicted Defective rate on Process A, RMSE Score

그림 16은 그림 15의 변수들을 기반으로 OLS 회귀 분석을 진행하여, 실제 불량률과 예측 불량률을 시각화한 것이고 전체 RMSE Score는 0.0117이다. 그래프의 x축은 데이터의 index를 의미하고, y축은 불량률을 의미한다. 주황색 그래프는 실제 데이터의 불량률을 나타내고, 파란색 그래프는 실제 데이터를 기반으로 불량률을 예측한 것이다. 분석 결과 실제 불량률과 예측한 불량률이 큰 차이가 없는 것을 확인할 수 있다.

OLS Regression Results						
Dep. Variable:		불량률	R-squared:	0.944		
Model:	Least Squares	Adj. R-squared:	0.942			
Method:	Least Squares	F-statistic:	540.9			
Date:	Fri, 25 Oct 2019	Prob (F-statistic):	5.52e-60			
Time:	02:26:27	Log-Likelihood:	114.35			
No. Observations:	99	AIC:	-222.7			
Df Residuals:	96	BIC:	-214.9			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t		
				[0.025		
				0.975]		
가열온도	0.7073	0.078	9.095	0.000	0.553	0.982
주입속도	0.2889	0.066	4.357	0.000	0.157	0.421
완성도	-0.8060	0.038	-21.475	0.000	-0.980	-0.731
Omniibus:	12.414	Durbin-Watson:	0.265			
Prob(Omniibus):	0.002	Jarque-Bera (JB):	13.289			
Skew:	0.878	Prob(JB):	0.00130			
Kurtosis:	3.368	Cond. No.	18.4			

그림 17. A 공정의 OLS 회귀 분석 결과표  
Fig. 17. OLS regression results table on Process A

그림 17은 OLS 회귀 분석 결과표이며, 이 중 R-squared는 주어진 데이터를 현재 모형이 얼마나 잘 설명하고 있는지를 나타내는 지수이다. 가열온도가 1 증가할 때 불량률에 미치는 영향은 +0.7073이고, 주입속도가 1 증가 할 때 불량률에 미치는 영향은 +0.2889이다. 이와 반대로, 제품완성도는 1 증가할 때 -0.8060의 영향을 미친다. window size가 1000개일 때, 1000개의 데이터의 가열온도, 주입속도의 평균이 1 증가하면 불량률은 각각 0.70, 0.28이 증가하고, 1000개 데이터 제품 완성도의 평균이 1씩 증가한다면, 불량률은 0.80 감소한다. 공정 B와 C도 같은 알고리즘을 적용하였다.

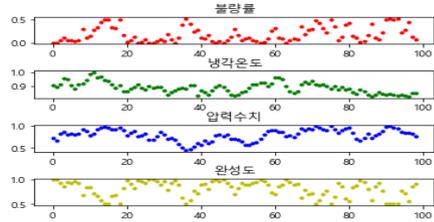


그림 18. B 공정 데이터의 window size별 평균값  
Fig. 18. Average value by window size of process B data

그림 18은 그림 15와 같은 방식을 B 공정에 적용한 변수들이다.

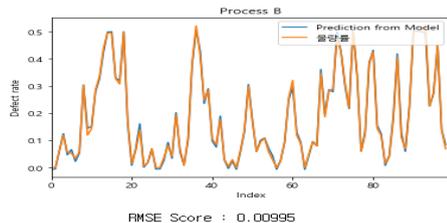


그림 19. B공정의 불량 예측 그래프와 RMSE 스코어  
Fig. 19. Predicted Defective rate on Process B, RMSE Score

그림 19는 그림 16과 같은 기법을 B 공정에서 실행한 그래프이고 전체 RMSE Score는 0.00995이다. 그림 16의 A 공정 분석 결과와 마찬가지로 실제 불량률과 예측한 불량률이 큰 차이가 없는 것을 확인할 수 있다.

OLS Regression Results						
Dep. Variable:		불량률	R-squared:	0.965		
Model:	Least Squares	Adj. R-squared:	0.964			
Method:	Least Squares	F-statistic:	877.7			
Date:	Fri, 25 Oct 2019	Prob (F-statistic):	1.29e-69			
Time:	02:16:16	Log-Likelihood:	163.39			
No. Observations:	99	AIC:	-320.8			
Df Residuals:	96	BIC:	-313.0			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t		
				[0.025		
				0.975]		
냉각온도	1.0209	0.066	18.156	0.000	0.909	1.133
압력수치	0.0484	0.042	1.163	0.248	-0.034	0.131
완성도	-0.9567	0.032	-29.570	0.000	-0.989	-0.973
Omniibus:	0.957	Durbin-Watson:	0.306			
Prob(Omniibus):	0.745	Jarque-Bera (JB):	0.723			
Skew:	-0.153	Prob(JB):	0.697			
Kurtosis:	2.713	Cond. No.	21.9			

그림 20. B 공정의 OLS 회귀 분석 결과표  
Fig. 20. OLS regression results table on Process B

그림 20은 OLS 회귀 분석 결과표이며, 냉각 온도, 압력 수치, 제품완성도 각각이 1씩 증가할 때마다 1.02, 0.04, -0.93의 영향을 미친다.

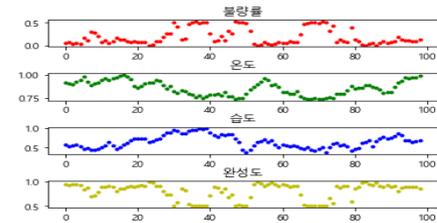


그림 21. C 공정 데이터의 window size별 평균값  
Fig. 21. Average value by window size of process C data

그림 21은 그림 15와 같은 방식을 C 공정에 적용한 변수들이다.

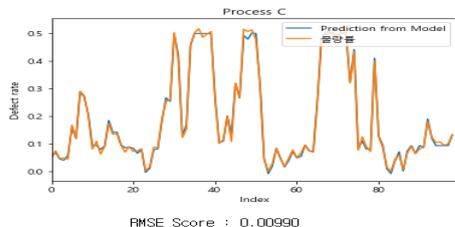


그림 22. C공정의 불량 예측 그래프와 RMSE 스코어  
Fig. 22. Predicted Defective rate on Process C, RMSE Score

그림 22는 그림 16과 같은 기법을 C 공정에서 실행한 그래프이고 전체 RMSE Score는 0.00990이다. 그림 16

과 그림 19의 A, B 공정 분석 결과와 마찬가지로 실제 불량률과 예측한 불량률이 큰 차이가 없는 것을 확인할 수 있다.

OLS Regression Results						
Dep. Variable:	불량률	R-squared:	0.903			
Model:	OLS	Adj. R-squared:	0.900			
Method:	Least Squares	F-statistic:	239.6			
Date:	Fri, 25 Oct 2019	Prob. (F-statistic):	1.30e-48			
Time:	02:22:40	Log-Likelihood:	-113.89			
No. Observations:	99	AIC:	-215.6			
Df Residuals:	96	BIC:	-236.0			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
온도	1.1489	0.072	15.973	0.000	1.006	1.292
습도	0.2074	0.046	4.528	0.000	0.116	0.298
완성도	-1.1622	0.062	-19.955	0.000	-1.285	-1.040
Omnibus:	11.977	Durbin-Watson:	0.185			
Prob(Omnibus):	0.003	Jarque-Bera (JB):	3.975			
Skew:	0.092	Prob(SB):	0.137			
Kurtosis:	2.036	Cond. No.	15.8			

그림 23. C 공정의 OLS 회귀 분석 결과표  
 Fig. 23. OLS regression results table on Process C

그림 23은 OLS 회귀 분석 결과표이며, 경화 온도, 경화 습도, 제품완성도 각각은 1씩 증가할 때마다 1.14, 0.20, -1.16의 영향을 미친다.

## V. 결론

본 논문에서는 데이터 생성기를 활용하여 스마트공장에서 발생할 수 있는 가상 데이터를 생성하고, 이를 활용하여 불량품의 발생 원인을 분석하고 설비의 교체주기를 예측하는 방법을 연구하였다. 본 연구를 통해 스마트 공장의 생산량 및 효율성을 향상시킬 수 있을 것으로 기대된다. 또한, 데이터 생성기를 이용하여 생성한 가상 데이터가 아닌 실제 공정에서 생성되는 데이터를 활용하여 분석한다면 더욱 유의미한 결과를 도출할 수 있을 것으로 기대된다.

## References

- [1] Smart Manufacturing Technologies and Standards, ETRI Insight, Jan 2018.
- [2] Smart Manufacturing R&D Roadmap, Ministry of Trade, Industry and Energy, Mar 2019.
- [3] Hyeong-Uk Park, "Trends in Production and Manufacturing Technologies in Smart Factory", The Journal of The Korean Institute of Communication Sciences, Vol. 33, No. 1, pp. 24-29, Dec 2015.
- [4] Sang-Do No, "Smart Factory and Cyber-Physical System Technology", The Journal of The Korean Institute of Communication Sciences, Vol. 33, No. 11, pp. 3-7, Oct 2016.

- [5] Dong-Jin Shin, Ji-Hun Park, Kwang-Jin Kwak, Jeong-Min Park, Jeong-Joon Kim, "Big Data-based Sensor Data Processing and Analysis for IoT Environment", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 19, No. 1, pp. 117-126, Feb 2019.  
DOI: <https://doi.org/10.7236/IIBC.2019.19.1.117>
- [6] Hyeon-Jong Lee, "Bigdata Hadoop Platform", The Journal of The Korean Institute of Communication Sciences, Vol. 29, No. 11, pp. 43-47, Oct 2012.
- [7] Seung-Yeon Hwang, Ji-Hun Park, Ha-Young Youn, Kwang-Jin Kwak, Jeong-Min Park, Jeong-Joon Kim, "Big Data based Medical Clinical Results Analysis", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 19, No. 1, pp. 187-195, Feb 2019.  
DOI: <https://doi.org/10.7236/IIBC.2019.19.1.187>
- [8] Dong-Ha Shin, Kwang-Ho Choi, Chang-Bok Kim, "Deep Learning Model for Prediction Rate Improvement of Stock Price Using RNN and LSTM", The Journal of Korean Institute of Information Technology, Vol. 15, No. 10, pp. 9-16, Oct 2017.  
DOI: <http://dx.doi.org/10.14801/jkitt.2017.15.10.9>
- [9] Heung-Seon Oh, Yuchul Jung, "Analysis of Massive Scholarly Keywords using Inverted-Index based Bottom-up Clustering", The Journal of the Korea Academia-Industrial cooperation Society, Vol. 19, No. 11, pp. 758-764, Nov 2018.  
DOI: <https://doi.org/10.5762/KAIS.2018.19.11.758>
- [10] Gi-Sung Kim, Ik-Rae Jeong, "Practical Privacy-Preserving DBSCAN Clustering Over Horizontally Partitioned Data", The Journal of the Korea Institute of Information Security and Crpytology, Vol. 20, No. 3, pp. 105-111, Jun 2010.
- [11] Uh Soo Kyun, Sung-Hoon Cho, Jeong-Joon Kim, Young-Gon Kim, "A Study on Perception for Public Safety of Seoul Citizens using Multiple Regression Analysis", The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 18, No. 1, pp. 195-201, Feb 2018.  
DOI: <https://doi.org/10.7236/IIBC.2018.18.1.195>

## 저 자 소 개

### 황 승 연(준회원)



• Seung-Yeon Hwang received his BS in Computer Engineering at Korea Polytechnic University in 2019. He is currently a Master's course in the department of Smart Manufacturing Engineering at Korea Polytechnic University. His research interests include Big Data, Internet of Things(IoT), Machine Learning, etc.

곽 경 민(준회원)



- Kyung-Min Gwak received his BS in Computer Engineering at Korea Polytechnic University in 2019. He is currently a Master's course in the department of Smart Manufacturing Engineering at Korea Polytechnic University. Recently, he is research on AI, Data-Mining technology.

신 동 진(준회원)



- Dong-Jin Shin received his BS in Engineering at Korea Polytechnic University in 2018. He is currently a Master's course in the department of Smart Manufacturing Engineering at Korea Polytechnic University. His research interests include Big Data, Internet of Things(IoT), Network Security, etc.

곽 광 진(준회원)



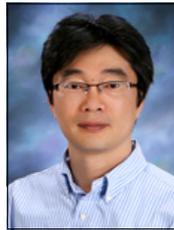
- Kwang-Jin Kwak received his MS in computer Science at Konkuk University in 2010 and 2016. He is currently a studying at Korea Polytech University for a doctor's course in Department of Smart Manufacturing Engineering. His intersts GIS, Information Retrieval, Text Mining, Database, NoSQL, etc.

노 영 주(정회원)



- Young J Rho recived his BS at Korea University, MS at FDU, Ph.D at UNSW. He has been a professor at the KPU since 2005. His interests are SW, HCI, IoT, ML

박 경 원(정회원)



- Kyungwon Park received his BS and MS in Mathematics at Chung-Ang University in 1995 and 1998, respectively. In 2004, he received his Ph. D in at University of South Carolina . He is currently a professor at the department of Computer Science at Korea Polytechnic University. His research interests include Nonlinear Approximation Theory, Image Analysis, Machine Learning, etc.

박 정 민(정회원)



- Jeong-Min Park received his BS in Computer Science at Korea Polytechnic University in 2003. He received his MS and PhD in at SungKyunKwan University in 2005 and 2009, respectively. He is currently a professor at the department of Computer Science at Korea Polytechnic University. His research interests include Cyber Physical System(CPS), Autonomic Computing, Software Engineering, etc.

김 정 준(정회원)



- Jeong Joon Kim received his BS and MS in Computer Science at Konkuk University in 2003 and 2005, respectively. In 2010, he received his PhD in at Konkuk University. He is currently a professor at the department of Computer Science at Korea Polytechnic University. His research interests include Database Systems, BigData, Semantic Web, Geographic Information Systems (GIS) and Ubiquitous Sensor Network (USN), etc.

※ 이 논문은 산업통상자원부의 재원으로 한국산업기술진흥원(KIAT)의 지원을 받아 수행된 연구임. (2019년 스마트공장 운영설계 전문 인력 양성사업, 과제번호 : N0002429)