

# 다양한 환경정보 데이터를 활용한 이상치 탐지 방법 연구

백민영\*, 최강인\*, 배성재\*, 김동훈\*, 조세운\*, 김병준\*, 정성환\*

\*한국전자기술연구원

e-mail:{bmy514, cki921, bsj940528, clickmiss123, swcho, jun0420, shjeong}@keti.re.kr

## A Study on Anomaly Detection Methods Using Various Environmental Data

Minyoung Baek\*, Kangin Choi\*, Seongjae Bae\*, Donghoon Kim\*,

Seowoon Cho\*, Byoungjun Kim\*, Sunghwan Jeong\*

\*Korea Electronics Technology Institute

### 요약

4차 산업혁명에 따라 농업 기술의 확산으로 작물 병해충 예측, 식생지수 분석, 생육환경 제어 최적화 등 다양한 연구 등이 진행되었지만, 스마트팜 시설 내 설치된 센서의 고장 및 통신 문제로 인한 데이터 손실 및 이상치 등과 같은 다양한 문제에 대해 이상치 탐지 가능한 솔루션 확보가 필요하다. 본 연구에서는 다양한 환경에서 수집된 공개데이터를 활용하여 이상치 탐지 방법에 대해 연구를 진행하였다.

## 1. 서론

작물을 효과적으로 생육 및 수확하는 것은 작물 생산성을 개선하고 품질을 향상시키는 중요한 수단으로, 현대 농업 생산에서 중요한 과제 중 하나이다. 일반적으로 국내 작물 생육 및 생산에서는 스마트팜 또는 비닐하우스 등 시설 공간에서 재배하고 있으며, 생육환경 최적화를 위해 온도, 습도, 이산화탄소, EC, 광량 등과 같은 다양한 센서를 활용하여 데이터를 수집하고 인공지능 기술 분석 또는 작업자의 경험적 판단으로 환경 제어를 하고 있다.

스마트팜 및 정밀농업 기술의 고도화로 스마트팜, 비닐하우스, 그리고 수직농장 등 다양한 시설 내에서 다양한 품종의 작물이 재배되며, RGB 및 초분광 카메라를 활용한 병해충 예측, 식생지수 분석 등이 활발히 이루어지고 있다. 그러나 생육환경 제어는 여전히 센서 데이터 기반의 수작업에 의존하며, 센서의 고장과 이상치로 인해 최적화에 한계가 있다.

최근 스마트팜 시설 내 센서를 활용하여 생육환경 최적화 및 제어를 위한 센서 고장 및 오작동에 의한 이상 상황 분석 시스템[1], 데이터 품질 향상을 위한 이상치 및 결측치 보정 기법[2], 온실형 환경 데이터를 다중회귀분석으로 학습한 센서 데이터 이상징후 탐지 시스템[3]에 관한 연구가 진행되고 있다. 한편, 본 연구는 실시간성을 고려한 슬라이딩 윈도우 기반 통계적 접근을 통한 센서 데이터의 이상치 탐지에 대한 방법을 제시한다.

## 2. 제안방법

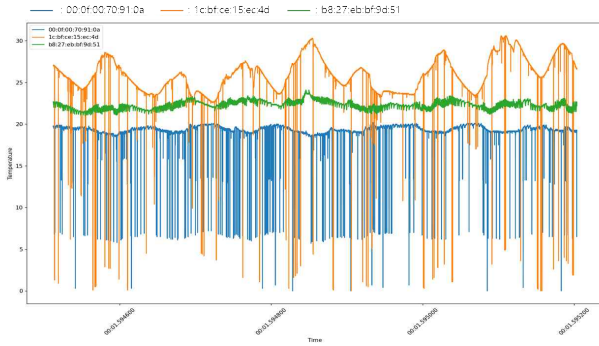
### 2.1 다양한 환경 정보 포함된 공개 데이터

일반적인 스마트팜 환경정보 데이터는 온도, 습도, 이산화탄소 농도, 일조량, 토양 수분, 산소, EC, 그리고 PH 등과 같은 다양한 데이터로 구성되어 있지만, 대부분 농가에서는 대표적인 환경 정보인 온도, 습도, 이산화탄소 농도 등 데이터를 활용한다. 본 연구에서는 대표적인 환경 정보인 온도, 습도 정보를 선정하여 이상치 탐지 연구를 진행하였으며, 다양한 환경에 설치된 센서를 통해 획득한 공개 데이터인 Environmental Sensor Telemetry Data를 활용하였다.

Environmental Sensor Telemetry Data는 공개 데이터로 8일간 3개의 센서 디바이스를 통해 온도, 습도, 일산화탄소, LPG 등 데이터가 포함되어 있다. 아래 표1은 수집된 센서에 대한 환경 특징을 나타내고, 그림1은 수집된 데이터에 대한 온도 정보를 가시화한 결과를 나타낸다. 이 중, 수직 농장의 환경적인 특징을 반영하여 안정적인 환경에서 측정된 2개의 센서 데이터를 가지고 연구를 진행하였다.

[표 1] Environmental Sensor Telemetry Data 환경 정보

센서 ID	환경특징
00:0f:00:70:91:0a	안정적인 환경이지만 낮은 온도와 높은 습도
1c:bf:ce:15:ec:4d	온도와 습도가 자주 변동되는 환경
b8:27:eb:bf:9d:51	안정적인 환경이지만 높은 온도와 낮은 습도

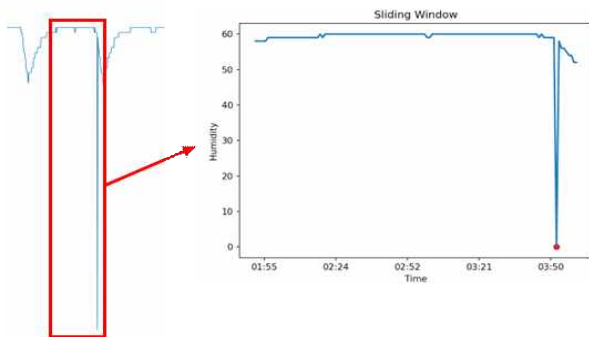


[그림 1] Environmental Sensor Telemetry 데이터 가시화

## 2.2 환경정보 센서 이상치 탐지 방법

본 연구에서는 실시간 환경정보 센서를 통해 수집되는 데이터를 기반으로 이상치 탐지 방법에 대해 연구하였다. 연구한 방법은 슬라이딩 윈도우 기반 통계적 방법으로, 시간적 흐름에 따라 수집된 환경 정보가 포함된 데이터 스트림을 고정된 크기의 윈도우 단위로 분할한 후 각 윈도우 내에서 평균과 표준편차 등의 통계량을 계산하여 이상치를 판정한다.

아래 그림2는 실시간으로 센서를 통해 수집되는 환경정보 데이터를 슬라이딩 윈도우 기반으로 이상치를 탐지하는 과정을 보여준다.

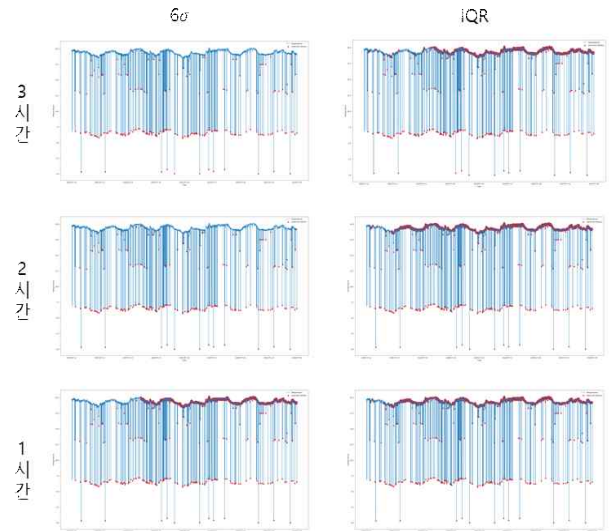


[그림 2] 수집된 환경정보 데이터 기반 이상치 탐지 과정

제안한 방법인 슬라이딩 윈도우 기반 통계적 방법에 대해 이상치 탐지의 적용 가능성을 확인하기 위해 본 연구에서는 IQR(InterQuartile Range)와 6시그마 방법을 비교하였다. IQR 방법은 전체 데이터 중 가운데 50%가 분포하는 구간의 길이를 제3사위분수(Q3)와 제1사위분수(Q1)로 계산하여 이상치를 탐지하는 방법으로, 표준편차와 달리 극단값의 영향을 상대적으로 적게 받으며, 정규분포와 같이 특정한 분포 형태를 가정하지 않으므로 다양한 유형의 데이터에 적용할 수 있다. 또한, 6시그마 방법은 전체 데이터의 평균( $\mu$ )과 표준편차( $\sigma$ )를 기준으로 상하한( $\mu \pm 6\sigma$ )을 설정하여 이 범위를 벗어난 값을 이상치로 탐지하는 방법으로, 평균값에서 크게 벗어난 극단값까지 포함하여 분석하기 때문에 센서 데이터의 전체 변동성을 고려한 이상치 탐지가 가능하다.

선정한 IQR 및 6시그마 방법을 통해 다양한 환경에서 수집된 공개 데이터를 활용하여 슬라이딩 윈도우 크기에 따른 이상치 탐지 적용을 비교하기 위해 슬라이딩 윈도우 크기를 1시간, 2시간, 그리고, 3시간 단위로 설정하여 실험을 진행하였다. 실험 결과, 센서 00:0f:00:70:91:0a의 온도 데이터에 대해 슬라이딩 윈도우 크기별 IQR과 6 $\sigma$ 를 사용한 이상치 탐지 결과는 아래 그림 3과 같으며, 센서별 온습도 데이터에 대해 탐지된 이상치 개수는 표 2와 같다.

센서 00:0f:00:70:91:0a

[그림 3] 슬라이딩 윈도우 크기별 IQR 및 6 $\sigma$  이상치 탐지 결과

[표 2] 탐지된 이상치 개수

윈도우 크기	00:00:0f:00:70:91:0a				b8:27:eb:bf:9d:51			
	IQR		6 $\sigma$		IQR		6 $\sigma$	
	온도	습도	온도	습도	온도	습도	온도	습도
1시간	92834	66219	74903	379	145614	114043	102402	1200
2시간	97608	77124	405	349	101479	18210	12	42
3시간	75770	5991	284	333	10053	16413	0	138

그림 3과 표 2를 통해 IQR 방법은 6시그마 방법에 비해 과도하게 이상치를 탐지된 것을 확인하였고, 슬라이딩 윈도우 크기가 1시간인 경우 다른 슬라이딩 윈도우 크기에 비해 탐지된 이상치 수가 더 많은 것을 확인하였다. 이를 통해 슬라이딩 윈도우 크기와 통계적 방법의 선정에 따라 환경정보 센서를 통해 획득한 데이터의 이상치 탐지 결과에 영향을 제공하는 것으로 판단된다.

## 3. 결론

본 논문에서는 스마트팜 시설 내 센서 데이터 이상치 탐지를 위하여 다양한 환경에서 수집된 공개 데이터를 활용하여 슬라이

딩 윈도우 기반 IQR 및 6시그마 통계적 방법을 활용하여 이상치 탐지에 관한 연구를 수행하였다. 제안한 방법을 통해 슬라이딩 윈도우 크기와 통계적 방법의 선정에 따라 환경정보 센서를 통해 획득한 데이터의 이상치 탐지 결과에 영향을 제공하는 것을 확인하였다. 이를 통해, 본 연구에서는 제안한 방법을 실시간 센서 데이터 모니터링의 기초자료로 활용할 예정이며, 향후 제안한 방법과 딥러닝 기술을 활용하여 이상치 및 결측치 복구에 대한 연구를 진행할 예정이다.

#### ACKNOWLEDGMENT

본 연구는 산업통상자원부와 한국산업기술진흥원의 “지역혁신 클러스터육성(R&D, P0025455)”사업의 지원을 받아 수행된 연구결과임

#### 참고문헌

- [1] 최휘민, 김주만, “스마트팜 ICT기기의 이상탐지 시스템”, 한국인터넷방송통신학회 논문지, 제 19권 2호, pp. 169–174, 2019년.
- [2] 이성재, 심현, “스마트팜 데이터 품질 향상을 위한 이상치 및 결측치 보정 방법에 관한 연구”, 한국전자통신학회 논문지, 제 19권 5호, pp. 1027–1034, 2024년.
- [3] 이철원, 안수용, 김재영, 안형태, “환경 데이터를 활용한 온실형 스마트팜에서 센서 이상 탐지 시스템”, 한국데이터정보과학회지, 제 32권 6호, pp. 1237–1248, 2021년.