

스마트팜 환경 데이터의 이상치 탐지를 통한 신뢰도 향상 기법 연구 - LOF 기반 접근

양광호, 이명훈*

*국립순천대학교

yg8686@naver.com, *leemh777@scnu.ac.kr

A Study on Improving the Reliability of Smart Farm Environmental Data Using LOF-Based Outlier Detection

Yang Kwang Ho, Lee Meong Hun*

*Sunchon National Univ.

요약

본 연구는 스마트팜 온실 환경에서 수집된 공공 데이터를 기반으로, LOF를 활용한 이상치 탐지 기법을 적용하여 환경 데이터의 신뢰도 향상 가능성을 분석하였다. 2021년 한 해 동안 측정된 7,382개의 시계열 환경 데이터를 전처리하고, LOF 알고리즘을 적용하여 이상치를 탐지하였다. 탐지된 이상치는 대부분 밀도가 낮거나 급격한 변화가 있는 지점에서 발생하였으며, 이는 센서 오류나 환경 변동성으로 인한 것으로 해석된다. 또한, 이상치와 정상 데이터를 변수별로 통계 비교한 결과, 외부 온도, 풍속, 일사량 등에서 유의한 차이가 나타났으며, 이는 LOF 기반 기법이 단순 임계값 필터링보다 복합적인 이상 패턴에 효과적으로 대응할 수 있음을 시사한다. 본 연구는 향후 스마트팜 제어 알고리즘에 적용 가능한 데이터 전처리 기반을 제공하며, 실시간 제어 시스템의 안정성과 신뢰도 제고에 기여할 수 있을 것으로 기대된다.

I. 서론

스마트팜은 ICT를 농업에 접목하여 작물 생육 환경을 자동으로 제어하고, 농업 생산성과 자원 효율을 극대화하는 첨단 농업 시스템이다. 특히 온실, 노지, 수경재배 등 다양한 재배 조건에서 센서를 통해 수집된 환경 데이터는 생육 예측, 자동 관수, 병해충 진단 등 다양한 응용에 활용되고 있다[1].

그러나 센서 오류, 통신 장애, 외부 간섭 등의 요인으로 인해 실제 측정값과 일치하지 않는 비정상적인 데이터, 즉 '이상치(Outlier)'가 발생할 수 있으며, 이러한 이상치는 모델의 예측 정확도를 저하시키고 제어 시스템의 신뢰성을 위협할 수 있다[1-3]. 따라서 수집된 환경 데이터의 신뢰도를 향상시키기 위해 이상치를 효과적으로 탐지하고 제거 또는 보정하는 기술이 필수적이다.

기존에는 단순한 임계값 기반 필터링이나 평균/표준편차 기준 탐지 기법이 널리 사용되어 왔지만, 이는 다변량 데이터에 대한 비선형적 이상 패턴을 놓칠 수 있다는 한계를 지닌다. 최근에는 머신러닝 기반 이상치 탐지 기법이 주목받고 있으며, 그 중 Local Outlier Factor(LOF)는 데이터의 밀도 기반 접근을 통해 국지적인 이상치를 효과적으로 식별할 수 있는 비지도 학습 기반 방법으로 알려져 있다 [4-5].

본 연구에서는 스마트팜 온실 환경에서 수집된 다변량 센서 데이터를 대상으로 LOF 기반 이상치 탐지 기법을 적용하고, 정상 데이터와 이상 데이터의 통계적 차이를 분석함으로써 환경 데이터의 신뢰도 향상 가능성을 검토하였다. 이를 통해 향후 스마트팜 제어 알고리즘에 적용 가능한 데이터 전처리 기반을 제시하고자 한다.

II. 데이터 구성 및 전처리 과정

본 연구에서는 스마트팜코리아에서 제공하는 공공 데이터 중 온실 환경 데이터를 활용하였다. 해당 데이터는 특정 스마트팜 온실에서 일정 기간 동안 자동 센서를 통해 수집된 자료로, 내부 온도(℃), 외부 온도(℃), 내부

습도(%), 이산화탄소 농도(ppm), 외부 풍속(m/s), 외부 일사량(W/m²) 등 10개의 주요 환경 변수들을 포함하고 있다.

데이터는 시계열 형태로 구성되어 있으며, 이상치 탐지 실험을 위해 2021년 2월부터 2021년 12월까지의 자료를 활용하였다. 전처리를 위해 원본 데이터에서 이상치 영향을 크게 받고 실질적인 생육에 영향을 줄 수 있는 6개의 환경센서 변수를 선택하였다. 총 7,382개의 샘플을 확보하고 결측값으로 인한 노이즈를 제거하여 분석에 적합한 형태로 변환하였다.

III. 이상치 탐지 방법: LOF

이상치 탐지를 위해 본 연구에서는 밀도 기반 이상치 탐지 기법인 LOF를 활용하였다. LOF 값이 1에 가까우면 정상, 1보다 크면 주변 밀도보다 낮아 이상치일 가능성이 높은 것으로 간주한다. 본 연구에서는 sklearn의 LocalOutlierFactor 모듈을 사용하였으며, 주요 설정은 다음과 같다:

- 이웃 수($n_{\text{neighbors}}$): 20
 - LOF 임계값: LOF 점수가 1.0 초과인 데이터를 이상치로 판단
- 해당 방법은 단순 임계값 기반 방식보다 다양한 변수의 복합적인 이상 징후를 파악할 수 있다는 점에서 스마트팜 환경 데이터 분석에 적합하다. 이상치 탐지 결과의 타당성을 평가하기 위해, 산점도나 그래프와 같은 방법으로 정상 데이터와 이상치 데이터를 시각적으로 비교하였다.
- 또한, 변수별 평균 및 표준편차 값을 산출하여, 이상치와 정상치 간의 통계적 차이를 분석하였다. 이를 통해 LOF 기반 이상치 탐지의 실효성을 정량적으로 검증할 수 있도록 하였다.

IV. 실험 결과

본 절에서는 LOF 기반 이상치 탐지 결과를 시각화하고, 이상치와 정상 데이터 간의 차이를 통계적으로 분석하였다.

그림 1은 내부 온도와 내부 습도를 x, y축으로 하여 LOF 기반 이상치 탐지 결과를 2차원 공간상에 표현한 것이다. 파란색 점은 정상으로 판단된 데이터이며, 빨간색 점은 이상치로 분류된 샘플이다.

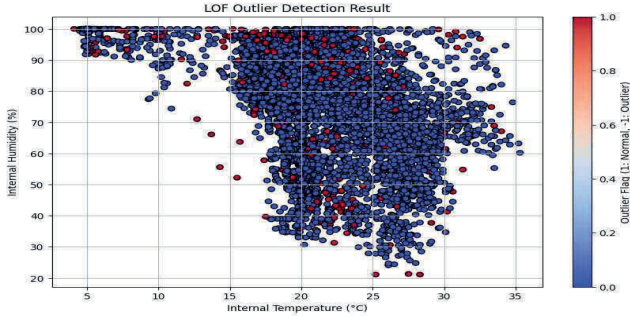


그림 1. LOF 기반 이상치 탐지 결과 시각화

Figure 1. Visualization of LOF-based outlier detection results

대부분의 데이터는 온도 20 - 30°C, 습도 60 - 95% 범위 내에 밀집되어 있으나, 경계 바깥 영역 또는 밀도가 낮은 영역에서 이상치가 집중적으로 탐지되었다. 이는 LOF가 지역 밀도 기반으로 이상치를 판별함을 반영하며, 기존의 단순 임계값 방식으로는 탐지하기 어려운 복합 이상 패턴까지 포착할 수 있음을 보여준다.

그림 2는 2021년 2월부터 12월까지의 내부 온도 시계열 데이터를 나타낸 것으로, 파란색 선은 전체 온도 변화, 빨간색 점은 LOF에 의해 이상치로 탐지된 지점을 나타낸다.

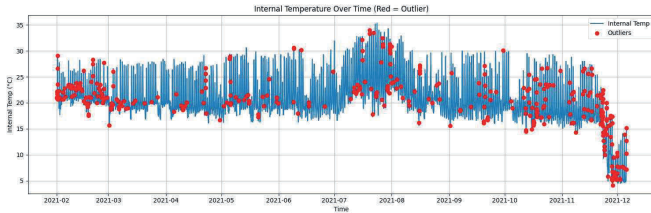


그림 2. 내부 온도 시계열 변화 및 이상치 시각화 결과

Figure 2. Visualization results of internal temperature time series changes and outliers

이상치는 특정 기간에 집중적으로 발생하기보다는 전반적으로 넓게 분포해 있으며, 특히 급격한 온도 변화가 나타나는 시점이나 낮은 온도 값에서 자주 발생하였다. 이는 센서 고장, 통신 오류, 외부 간섭 등 다양한 요인으로 인해 실측값과 불일치하는 데이터가 생성되었음을 시사한다.

이상치와 정상 데이터에 대해 각 변수의 평균과 표준편차를 산출하여 정량적으로 비교한 결과는 표 1과 같다.

표 1. 변수별 이상치와 정상 데이터 간 통계 지표 비교
Table 1. Comparison of statistical indicators between outliers and normal data by variable

구분	외부온도 (°C)	외부풍속 (m/s)	일사량 (W/m ²)	내부습도 (%)	CO ₂ 농도 (ppm)
정상 평균	13.62	1.29	235.67	83.62	496.58
이상치 평균	8.57	1.38	138.74	85.12	500.02
정상 표준편차	11.26	1.05	252.49	14.81	132.04
이상치 표준편차	13.62	1.32	222.08	15.28	142.75

이러한 결과는 LOF 탐지 결과가 단순히 값의 크기만으로 분류된 것이 아니라, 다변량 밀도 차이를 기반으로 다양한 패턴의 이상치를 탐지했음을 의미한다.

V. 결론

본 연구에서는 스마트팜 코리아에서 제공하는 공공 온실 환경 데이터를 활용하여, 밀도 기반 이상치 탐지 기법인 LOF를 적용함으로써 스마트팜 환경 데이터의 신뢰도 향상 가능성을 실증적으로 분석하였다.

전체 7,382개의 시계열 환경 데이터를 대상으로 LOF 점수를 계산하였으며, LOF 값이 1.0을 초과하는 데이터를 이상치로 정의하였다. 분석 결과, 내부 온도와 습도를 기준으로 한 2차원 산점도에서 이상치는 경계 바깥이나 국지적으로 밀도가 낮은 지점에 집중되는 양상을 보였다. 시계열 분석에서는 급격한 온도 변화나 비정상적인 패턴 구간에서 이상치가 주로 탐지되었으며, 이는 실제 현장에서 발생 가능한 센서 고장, 통신 오류, 외부 간섭 등의 영향을 반영하는 것으로 해석된다.

또한, 이상치와 정상 데이터를 변수별로 분류하여 평균 및 표준편차를 비교한 결과, 외부 온도, 일사량, 풍속, 습도, CO₂ 농도 등에서 통계적으로 유의한 차이가 나타났다. 이는 LOF가 단순 값의 크기가 아닌, 다변량 밀도 기반의 접근을 통해 복합적인 이상 패턴을 효과적으로 탐지할 수 있음을 시사한다.

이러한 결과는 스마트팜 제어 시스템의 데이터 전처리 단계에 LOF 기반 이상치 탐지 기법을 도입함으로써, 센서 데이터의 신뢰성을 높이고, 자동화 제어의 안정성과 정확도를 향상시킬 수 있는 가능성을 제시한다. 향후 연구에서는 LOF 외의 다른 이상치 탐지 알고리즘과의 비교 및 실시간 적용성 평가를 수행할 예정이다.

ACKNOWLEDGMENT

본 결과물은 농림축산식품부 및 과학기술정보통신부, 농촌진흥청의 재원으로 농림식품기술기획평가원과 재단법인 스마트팜연구개발사업단의 스마트팜다부처패기 지혁신기술개발사업의 지원을 받아 연구되었음(RS-2025-02219360)

참 고 문 헌

- [1] 이철원, 안수용, 김재영, 안형태, “환경 데이터를 활용한 온실형 스마트팜에서 센서 이상 탐지 시스템”, 한국데이터정보과학회지, 32(6), 1237-1248, 2021.
- [2] 이성재, 심현, “스마트팜 데이터 품질 향상을 위한 이상치 및 결측치 보정 방법에 관한 연구”, 한국전자통신학회 논문지, 19(05), 1027-1034, 2024.
- [3] Munir, M., Siddiqui, S. A., Dengel, A., & Ahmed, S. “DeepAnT: A deep learning approach for unsupervised anomaly detection in time series”, IEEE Access, 7, 1991 - 2005, 2019.
- [4] 김윤성, 문채영, 이상돈, 조민준, 신연순, “머신러닝 기반 멀티 IoT 센서를 사용한 대학 캠퍼스의 재실 감지 및 스마트 에너지 절감 솔루션” 한국정보기술학회 2024년도 추계종합학술대회 및 대학생논문경진대회, 959-963, 2024
- [5] 이용찬, 권용범, 이희관, “LOF 이상치 알고리즘을 활용한 국내 초미세먼지 발생 특성 탐지 분석”, 한국대기환경학회지, 37(1), 125-143, 2021.