

LSTM Autoencoder 기반 스마트 온실 센서 데이터 이상 탐지 구조 설계

정주원, 여현*

*국립순천대학교

fock9320@daum.net, yhyun@scnu.ac.kr

Design of an LSTM Autoencoder-based Anomaly Detection Architecture for Smart Greenhouse Sensor Data

Jeong Joo Won, Yoe Hyun*

*Sunchon National Univ

요약

스마트 온실에서 센서 데이터의 안정적 확보는 작물 생육과 제어 시스템의 신뢰성 유지에 중요한 요소로 작용한다. 본 연구는 LSTM Autoencoder 기반 구조를 설계하여, 정상 데이터를 학습하고 재구성 오차를 활용함으로써 비정상 데이터를 실시간으로 탐지할 수 있는 이상 탐지 체계를 제시한다. 제안된 구조는 데이터 수집, 모델 학습, 전처리, 이상 탐지, 결과 처리의 단계로 구성되며, 결측치 보정과 정규화, 시계열 윈도우 생성을 포함한 전처리 과정을 통해 데이터 품질을 보장한다. 학습 단계에서는 정상 데이터 분포를 반영해 모델을 구축하고, 실시간 단계에서는 입력 데이터와 복원 데이터 간의 차이를 기반으로 이상 여부를 판정한다. 이러한 구조는 기존 임계값 기반 방식의 한계를 보완하고, 스마트 온실 내 자동화 수준 향상과 안정적 생육 환경 운영 안정성 향상에 이바지할 것으로 기대된다.

I. 서론

스마트 온실은 온도, 습도, 이산화탄소(CO_2), 전기전도도(EC) 등 다양한 환경 인자를 자동으로 제어하여 작물 생육을 최적화하고 노동력을 절감하는 차세대 농업 시스템으로 주목받고 있다. 그러나 실제 운영 과정에서 센서 고장, 통신 오류, 급격한 환경 변화로 인해 비정상 데이터가 발생하며, 이로 인해 제어 시스템의 신뢰성이 약화되고, 작물 생육 환경의 안정성이 흔들릴 수 있다. 이 문제를 극복하기 위해서는 센서 데이터의 이상을 즉시 탐지할 수 있는 기술이 필요하다. 기존의 임계값 기반 이상 탐지 기법은 환경 변화에 따라 값을 재설정해야 하는 번거로움이 있으며, 통계 기반 접근은 다변량 센서 데이터의 복잡한 상관관계를 충분히 반영하지 못한다는 한계가 있다. 이와 같은 한계를 보완하기 위해 최근에는 시계열 패턴 학습에 강점을 가진 딥러닝 기법이 주목되고 있으며, 그중 LSTM(Long Short-Term Memory)은 장기 의존성을 처리하여 시계열 데이터의 특징을 정밀하게 반영할 수 있어 응용 가치가 크다. LSTM Autoencoder는 정상 데이터의 분포를 학습한 후 입력을 복원하는 과정에서 발생하는 재구성 오차를 활용하여 이상 여부를 판단하는 방식으로 동작한다. 정상 데이터는 낮은 오차를 보이지만, 비정상 데이터는 높은 오차를 나타내므로 효과적인 이상 탐지가 가능하다. 이러한 특성은 스마트 온실 환경에서 센서 데이터의 안정성과 신뢰성을 확보하는 데 적합하다.

본 연구에서는 LSTM Autoencoder 기반의 구조를 설계하여 스마트 온실 센서 데이터의 이상 탐지를 구현하고자 한다. 제안된 구조는 다변량 시계열 데이터를 입력으로 정상 패턴을 학습하고, 재구성 오차를 기반으로 실시간 이상을 탐지하는 과정을 포함한다. 이를 통해 온실 운영의 자동화 수준을 높이고, 안정적인 생육 환경 유지에 기여할 수 있을 것으로 기대된다. 본 논문의 구성은 다음과 같다. II장에서는 관련 연구를 검토하고, III장에서는 제안하는 구조 설계와 처리 흐름을 제시한다. 마지막으로 IV장에서는 기대 효과와 향후 연구 방향을 논의한다.

II. 시스템 개요

본 연구에서 제안하는 이상 탐지 구조는 스마트 온실 내부에서 수집되는 센서 데이터를 기반으로 동작하며, 전체 흐름은 데이터 수집, 전처리, 모델

학습, 이상 탐지, 결과 처리의 단계로 구성된다. 온실 내부에는 온도, 습도, 이산화탄소(CO_2), 전기전도도(EC) 등 주요 환경 인자를 측정하는 센서가 배치되어 있으며, 각 센서로부터 실시간으로 데이터가 수집된다. 수집된 데이터는 네트워크를 통해 중앙 서버 또는 클라우드에 저장되며, 이후 LSTM Autoencoder 모델의 입력값으로 사용된다.

시스템은 정상 상태의 센서 데이터를 기반으로 모델을 학습하여 정상 분포를 표현하는 잠재 벡터를 생성한다. 새로운 데이터가 입력되면 모델은 이를 복원하는 과정을 거치고, 입력 데이터와 복원 데이터 간의 재구성 오차를 계산한다. 오차 값이 정상 범위를 벗어날 경우 해당 데이터를 이상치로 판정한다. 이상 탐지 결과는 시각화 모듈을 통해 관리자가 직관적으로 확인할 수 있으며, 필요 시 자동 제어 시스템과 연계하여 환기, 냉난방, 관수 장치의 동작을 보조하도록 설계할 수 있다.

이와 같은 시스템 개요는 스마트 온실 운영의 안정성을 강화하고, 센서 이상이나 환경 급변 상황을 조기에 인식하여 피해를 최소화할 수 있다는 점에서 의의가 있다.

II-I. 데이터 전처리

스마트 온실 센서 데이터는 장기간 연속적으로 수집되며, 측정 환경이나 장치 상태에 따라 결측치나 이상값이 발생할 수 있다. 따라서 모델 학습 이전에 전처리 과정을 수행하여 데이터 품질을 확보하는 것이 중요하다. 첫째, 결측치 처리는 센서 오류나 통신 장애로 인해 누락된 값을 보간법(interpolation) 또는 인접 구간 평균값 대체 방법을 통해 보정한다. 필요 시 비정상적으로 장시간 누락된 데이터는 학습에서 제외한다.

둘째, 정규화(Normalization) 과정은 센서 종류별 측정 범위가 상이하기 때문에 모든 데이터를 동일한 스케일로 맞추기 위해 수행된다. 본 연구에서는 Min-Max Scaling 기법을 적용하여 0과 1 사이 값으로 변환하였다. 이를 통해 모델이 특정 센서 값에 편향되지 않고 학습할 수 있도록 한다. 셋째, 시계열 윈도우 구성은 연속된 데이터를 일정한 구간 단위로 묶어 모델에 입력하는 절차이다. 예를 들어, 10분 간격으로 측정된 데이터를 5개씩 연결하여 하나의 시퀀스로 구성함으로써 LSTM Autoencoder가 시간적 패턴을 학습할 수 있도록 한다. 이 과정은 센서별로 동일하게 적용되며, 결과적으로 입력 데이터는 샘플 수 \times 시퀀스 길이 \times 특징 수(센서 수)의

3차원 형태로 변환된다.

넷째, 데이터 분리는 학습과 검증의 일반화를 위해 필수적이다. 수집된 정상 데이터를 학습용(70%)과 검증용(30%)으로 분리하여 모델의 학습 성능과 재구성 오차 안정성을 평가할 수 있도록 한다.

이러한 전처리 단계를 거친 후 모델에 입력되는 데이터는 잡음이 최소화된 형태로 정제되며, LSTM Autoencoder가 정상 패턴을 보다 안정적인 학습을 가능하게 하는 기반이 마련된다.

항목	임계값 기반 탐지	LSTM Autoencoder
데이터 처리 방식	센서값을 기준 범위와 단순 비교	시계열 정상 패턴 학습 후 복원 오차 기반 탐지
장점	구현이 간단, 즉시 적용 가능	다변량 시계열 반영, 작은 이상·복합 이상 탐지 가능
한계	계절·환경 변화 따라 기준 재설정 필요	학습 및 연산 자원 필요, 데이터 해상도에 성능 의존

표 1. 임계값 기반 탐지와 LSTM Autoencoder 기반 탐지 기법의 비교
Table 1. Comparison between Threshold-based Detection and LSTM Autoencoder-based Detection Methods

II-II. LSTM Autoencoder 모델 구조

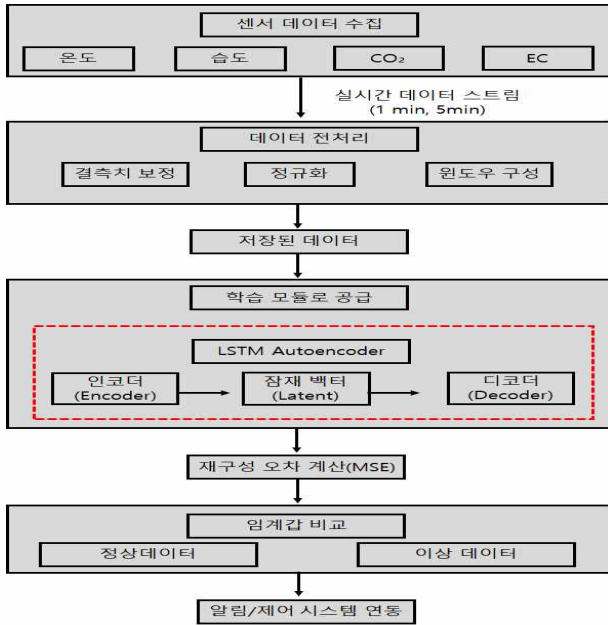


그림 1. LSTM Autoencoder 기반 스마트 온실 센서 데이터 이상 탐지 구조
Fig. 1. Architecture of Smart Greenhouse Sensor Data Anomaly Detection based on LSTM Autoencoder

LSTM Autoencoder는 시계열 데이터를 입력으로 받아 정상 패턴을 학습하고 이를 복원하는 과정을 통해 이상 탐지를 수행한다. 모델은 크게 인코더(Encoder), 잠재 벡터(Latent vector), 디코더(Decoder)로 구성된다. 인코더는 다변량 센서 데이터를 입력받아 시계열의 장기 의존성을 학습하고, 이를 저차원의 잠재 벡터로 압축한다. 이 과정에서 정상 데이터의 주요 패턴과 분포가 함축적으로 표현된다. 디코더는 인코더로부터 전달된 잠재 벡터를 다시 원래의 시계열 형태로 복원한다. 학습 과정에서는 입력과 출력의 차이가 최소화되도록 가중치가 조정된다. 손실 함수는 입력값과 복원값의 차이를 계산하는 재구성 오차(Reconstruction Error)로 정의된다. 본 연구에서는 평균제곱오차(Mean Squared Error, MSE)를 사용하여 학습을 진행하였다.

이 모델은 정상 데이터만을 학습하기 때문에 새로운 데이터가 입력되었을 때 정상

분포에서 벗어난 경우, 복원 과정에서 큰 오차가 발생하게 된다. 이 특성을 활용하여 이상 여부를 판단할 수 있다.

II-III. 이상 탐지 로직

이상 탐지는 입력 시퀀스와 복원 시퀀스 간의 재구성 오차를 계산한 뒤, 임계값(Threshold)을 초과하는 경우를 이상치로 판정하는 방식으로 수행된다. 먼저 각 시퀀스에 대해 입력값과 출력값 간의 차이를 산출한다.

학습 과정에서 얻어진 정상 데이터 오차 분포를 기반으로 90% 또는 95% 분위수를 임계값으로 설정한다. 이는 소수의 잡음 데이터를 고려하지 않으면서도 정상과 이상을 구분할 수 있는 유연한 기준을 제공한다.

입력 데이터가 임계값을 초과하면 이상으로 판정되며, 결과는 히트맵 및 시계열 그래프 형태로 시각화된다. 이를 통해 관리자는 특정 센서에서 발생한 이상 상황을 직관적으로 확인할 수 있다.

또한, 잠재 벡터를 활용한 탐지를 함께 적용하여 미세한 변화를 조기에 포착할 수 있도록 하였으며, 데이터 수집 간격에 따라 성능이 달라질 수 있음을 고려하여 설계하였다.

III. 결론

본 연구에서는 스마트 온실 환경에서 발생할 수 있는 센서 데이터의 이상을 탐지하기 위해 LSTM Autoencoder 기반 구조를 설계하였다. 제안된 구조는 다변량 센서 데이터를 입력으로 정상 패턴을 학습하고, 재구성 오차를 활용하여 이상 여부를 판단하는 방식으로 동작한다. 이를 통해 기존 임계값 기반 방식의 주관적 기준 설정 문제와 통계적 기준 방법의 한계를 개선할 수 있음을 확인하였다. 데이터 전처리 과정을 통해 결측치 보정, 정규화, 시계열 윈도우 구성 등을 수행함으로써 모델 학습의 안정성을 확보하였으며, LSTM Autoencoder는 정상 데이터의 분포를 잠재 벡터에 함축하여 복원 성능을 높였다. 또한, 재구성 오차 기반 임계값 설정과 시각화 기법을 통해 이상 탐지 결과를 직관적으로 제시할 수 있었다. 제안된 구조는 실제 스마트 온실 운영 환경에서 센서 고장이나 통신 장애와 같은 이상 상황을 조기에 탐지하고, 이를 자동 제어 시스템과 연계함으로써 작물 생육 안정성을 높일 수 있는 가능성을 지닌다. 더 나아가 습도, CO₂, EC 등 다양한 센서 데이터와의 융합, 영상 기반 데이터와의 확장 적용을 통해 멀티모달 이상 탐지 시스템으로 발전시킬 수 있다. 향후 연구에서는 엣지 컴퓨팅 환경에서의 실시간 배포와 다양한 데이터 해상도 조건에서의 성능 비교, 잠재 벡터 기반 탐지 기법의 고도화를 통해 보다 안정적이고 확장성 있는 스마트 온실 이상 탐지 구조를 구현할 필요가 있다.

ACKNOWLEDGMENT

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성 (Grand ICT연구센터) 사업의 연구결과로 수행되었음” IITP-2025-2020-0-01489

참 고 문 헌

- [1] 이정호, 임동혁, 김태현, 김만중, 박성진, 양오석, 백정현. (2023). LSTM Autoencoder를 활용한 스마트 온실 데이터 이상 탐지 구현. Journal of Knowledge Information Technology and Systems, 18(3), 587-596.
- [2] 정서현, (추가 저자명 미상). (2025). LSTM Autoencoder의 잠재 벡터를 사용한 자기상관 공정 모니터링. 학술대회 논문집, 2025..
- [3] 박준성, 이정섭, 유동건, 이충호, 허태욱, 이상금. (2025). LSTM Autoencoder와 Unsupervised Anomaly Detection 모델의 시간 해상도별 이상치 탐지 성능 분석. 대한전자공학회 하계학술대회 논문집, 2025..
- [4] 지평진, 이재현. (2024). LSTM Autoencoder를 이용한 자기상관 공정의 모니터링 절차. 응용통계연구, 37(2), 191-207.