

Isolation Forest을 활용한 축산 온·습도 이상치 탐지에 관한 연구

정광훈, 여현*

*국립순천대학교

gwanghoon5035@gmail.com, *yhyun@scnu.ac.kr

A Study on the Detection of Abnormalities in Livestock Temperature and Humidity Using Isolation Forest

Jung Gwang Hoon, Yoe Hyun*

*Sunchon National Univ.

요약

스마트 축사 분야에서 가축의 복지와 생산성은 온도, 습도 등 환경요인에 크게 영향을 받는 가장 주요한 관심사이며 이러한 환경요소를 정확히 모니터링하면 가축에 대한 필수적인 통찰력을 제공하여 가축 관리 결정에 직접적인 영향을 미칠 수 있다. 현재 지구온난화로 인해 평균 기온이 상승하면서 축산업 피해가 증가하고 있으며, 가축에 있어 고온 스트레스는 증체량 감소, 면역, 번식장애, 유량 저하 등 생산성에 부정적인 영향을 미치며 앞으로의 극한 기후 속에서 축산업을 지속하고 동물 복지를 실현하기 위해서는 가축 환경에 대한 이해를 바탕으로 적절한 대처 방안을 마련하는 것이 필요하다. 본 논문에서는 Isolation Forest을 활용한 축산 온·습도 이상치 탐지에 관한 연구를 진행하였으며 비지도 학습 방법인 Isolation Forest 알고리즘을 적용하여 축산에서 수집된 온도 및 습도 데이터의 이상값을 감지한다. 본 논문에서는 축산 온·습도 이상치 탐지를 위해 수집된 데이터를 로드, 이동 평균 및 시간데이터 생성, 모델 훈련 및 이상치 예측, 정밀도 성능 지표 및 이상치 출력을 통해 질병 조기 감지, 스트레스 감지, 가축 생산성 향상, 최적의 성장 조건 등에 기여할 것으로 기대된다.

I. 서론

스마트 축사 분야에서 가축의 복지와 생산성은 온도, 습도 등 환경요인에 크게 영향을 받는 가장 주요한 관심사이다[1].

이러한 환경요소를 정확히 모니터링하면 가축에 대한 필수적인 통찰력을 제공하여 가축 관리 결정에 직접적인 영향을 미칠 수 있다[2].

그러나 문제는 이러한 환경 변수를 모니터링하는 것뿐만 아니라 잠재적인 건강 위협이나 최적의 조건이 아닌 것을 나타낼 수 있는 이상 현상이나 특이점을 효율적으로 식별하는 것이다.[3].

환경 매개변수를 모니터링하고 분석하는 기존 방법은 임계값 기반 경고 또는 기본 통계 방법에 의존하는 경우가 많으며, 이는 유용하기는 하지만 새로운 문제를 나타내는 복잡한 패턴이나 미묘한 차이를 적절하게 포착하지 못할 수 있다.

결과적으로 농업 환경에서 수집된 환경 데이터의 감지 및 해석을 향상시키기 위해 기계 학습 기술을 활용하는 데 대한 관심이 높아지고 있다.

현재 지구온난화로 인해 평균 기온이 상승하면서 축산업 피해가 증가하고 있으며, 가축에 있어 고온 스트레스는 증체량 감소, 면역, 번식장애, 유량 저하 등 생산성에 부정적인 영향을 미치며 앞으로의 극한 기후 속에서 축산업을 지속하고 동물 복지를 실현하기 위해서는 가축 환경에 대한 이해를 바탕으로 적절한 대처 방안을 마련하는 것이 필요하다[4][5].

이러한 문제를 해결하기 본 논문에서는 Isolation Forest을 활용한 축산 온·습도 이상치 탐지에 관한 연구를 진행하였다. 현재 많은 연구에서 인공지능을 활용하고 있으며 비지도 학습 방법인 Isolation Forest 알고리즘을 적용하여 축산에서 수집된 온도 및 습도 데이터의 이상값을 감지한다. 본 논문에서는 축산 온·습도 이상치 탐지를 위해 수집된 데이터를 로드, 이동 평균 및 시간데이터 생성, 모델 훈련 및 이상치 예측, 정밀도 성능 지표 및 이상치 출력 순으로 설명하며 마지막으로 결론과 기대효과를 제시하고자 한다.

II. 본론

본 논문에서는 전남에 위치한 한 축산 농가에서 2023.01.01.~2023.10.31.까지 수집한 13,652개의 온·습도 데이터를 사용하여 연구를 진행하였다.

표 1. 데이터 특성

Table 1. Data Properties

특성	설명
INV_DATETIME	기록된 날짜와 시간
INT_TMP	내부 온도
INT_REL_HD	내부 습도
hour	INV_DATETIME에서 추출한 시간 데이터
day_of_week	INV_DATETIME에서 추출한 요일 데이터
day	INV_DATETIME에서 추출한 일 데이터
month	INV_DATETIME에서 추출한 월 데이터
year	INV_DATETIME에서 추출한 년도 데이터
rolling_avg_tmp	온도의 이동평균
rolling_avg_humidity	습도의 이동평균

<표 1>은 데이터 이상을 탐지하기 위하여 사용한 데이터 세트이다. 원시데이터 수집 및 품질 관리에 있어 1시간 단위 기준으로 학습 데이터 구축을 추진하였다.

```
# 데이터 로드 함수
def load_data(file_path):
    data = pd.read_csv(file_path)
    data['INV_DATETIME'] = pd.to_datetime(data['INV_DATETIME'])
    return data

# 이동 평균 및 시간 데이터 생성 함수
def preprocess_data(data):
    data['hour'] = data['INV_DATETIME'].dt.hour
    data['day_of_week'] = data['INV_DATETIME'].dt.dayofweek
    data['day'] = data['INV_DATETIME'].dt.day
    data['month'] = data['INV_DATETIME'].dt.month
    data['year'] = data['INV_DATETIME'].dt.year

    data['rolling_avg_tmp'] = data['INT_TMP'].rolling(window=3).mean()
    data['rolling_avg_humidity'] = data['INT_REL_HD'].rolling(window=3).mean()
    data.dropna(inplace=True)

    return data
```

그림 1. 데이터 로드 및 이동 평균 및 시간 데이터 생성
fig 1. Data Load and Move Average and Time Data Generation

머신러닝을 활용하여 수집된 데이터를 로드하고 이동 평균 및 시간 데이터를 생성하는 함수는 <그림 1>과 같다. <그림 1>은 load_data 함수를 사용하여 주어진 파일 경로에서 CSV 파일을 읽어 데이터를 로드하고 특정 컬럼의 날짜 및 시간 데이터를 분해 분석하여 원하는 형태로 조립하였으며 preprocess_data 함수를 사용하여 로드된 데이터에 시간 관련 새로운 특성을 추가하고, 일정한 크기의 윈도우를 사용하여 이동 평균을 계산하였다.

```
# 모델 훈련 및 이상치 예측 함수
def train_and_predict(data):
    features = data[['INT_TMP', 'INT_REL_HD', 'hour', 'day_of_week', 'day', 'month', 'year',
                    'rolling_avg_tmp', 'rolling_avg_humidity']]
    X = features.values

    X_train, X_test, y_train, y_test = train_test_split(features, range(len(features)), test_size=0.2, random_state=42)

    model = IsolationForest(n_estimators=100, contamination=0.01, random_state=42)
    model.fit(X_train)
    predictions = model.predict(X_test)
    anomaly_scores = model.decision_function(X_test)

    return X_test, y_test, predictions, anomaly_scores
```

그림 2. 모델 훈련 및 이상치 예측 함수
fig 2. Model training and outlier prediction functions

<그림 2>는 모델 훈련 및 이상치 예측 함수를 나타낸 것이다. 이상치를 식별하기 위해 IsolationForest 모델을 훈련하고 예측을 수행하는 함수를 사용하였으며 데이터 특성 선택, 데이터 분할, 모델 초기화 및 훈련, 이상치 예측 및 점수 계산을 진행하였다.

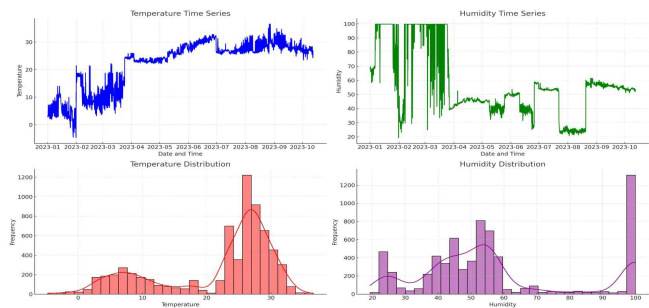


그림 3. 시간에 따른 온·습도 시각화
fig 3. Visualization and humidity according to time

<그림 3>은 시간에 따른 온·습도를 시각화해서 표현한 것이다. 왼쪽 상단의 그래프의 경우 온도 데이터를 시간 함수로 표현하였으며 눈에 띄는 변동과 함께 몇 달 동안의 온도 변화를 나타내며, 오른쪽 상단의 그림 같

은 경우 같은 기간 동안의 습도의 변화량을 나타내며 급격한 변동과 극적인 하락을 보이며 특히 4월과 8월에서 눈에 띄는 것을 확인할 수 있다. 왼쪽 하단의 그래프 같은 경우 해당 기간 동안 기록된 온도 분포를 보여주며 섭씨 20도 부근에서 최고점을 갖는 대략적인 정규 분포를 따르는 모습이 고 오른쪽 하단의 그래프 경우 습도의 분포를 자세히 보여준다.

날짜 및 시간	온도 (°C)	습도 (%)	이상치 점수
2023-01-01 06:59:00	3.04	68.83	-0.001275
2023-01-30 19:59:00	0.04	69.05	-0.026313
2023-01-29 00:59:00	-0.48	99.73	-0.027626
2023-01-27 23:59:00	-2.74	80.23	-0.027298
2023-01-29 22:59:00	2.21	98.08	-0.018282

그림 4. Isolation Forest 활용한 이상치 점수 계산
fig 4. Calculation of outlier scores using Isolation Forest

<그림 4>는 Isolation Forest 활용하여 이상치 점수 계산 나타낸 것이다. Isolation Forest 모델은 평균 경로 길이를 사용하여 이상 점수를 계산 시도하는데 경로 길이가 짧을수록 변칙 점수가 높아지고 경로가 길수록 변칙 점수가 낮아지는데(0에 가까워질수록) 이는 해당 값이 더 정상 적임을 나타낸다. 따라서 이상치 점수를 보고 0에서 멀어질수록 이상 현상으로 간주되는 지점을 파악할 수 있다.

III. 결론

본 논문에서는 Isolation Forest을 활용한 축산 온·습도 이상치 탐지에 관한 연구를 진행하였다. 비지도 학습 방법인 Isolation Forest 알고리즘을 적용하여 축산에서 수집된 온도 및 습도 데이터의 이상값을 감지한다. 본 논문에서는 축산 온·습도 이상치 탐지를 위해 수집된 데이터를 로드, 이동 평균 및 시간데이터 생성, 모델 훈련 및 이상치 예측, 정밀도 성능 지표 및 이상치 출력 순으로 설명하였다. 가축의 복지와 생산성은 온도, 습도 등 환경요인에 크게 영향을 받기때문에 인공지능을 활용하면 더 쉽고 빠르게 이상치를 탐지할 수 있으며 이러한 연구를 통해 질병 조기 감지, 스트레스 감지, 가축 생산성 향상, 최적의 성장 조건 등에 기여할 것으로 기대되며 추후 센서 데이터의 이상 탐지만 뿐만 아니라 다양한 구동기의 전력 데이터를 수집하여 이상 전력 사용량 등을 판단할 수 있도록 분석할 예정이다.

ACKNOWLEDGMENT

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음” (RS-2024-00259703)

참 고 문 헌

[1] 이동규, 이봉국, 정재영, 은지숙, and 김대희, “실시간 영상 처리 기술을 이용한 가축 이상징후 감지,” 디지털콘텐츠학회논문지, vol. 24, no. 10, pp. 2579-2584, 2023. (10.9728/dcs.2023.24.10.2579)

[2] 박태현 and 정재영, “가축 질병 확산 방지를 위한 딥러닝 기반 위험

요소 인식 시스템," 한국통신학회 인공지능 학술대회 논문집, pp. 334-335, 제주, 2023-09-20.

- [3] 이여름, 김남기, and 왕지남, "Isolation Forest를 활용한 사출 공정 중 이상 탐지," 한국생산제조학회 학술발표대회 논문집, pp. 409-409, 제주, 2023-12-13.
- [4] 유영대, 양광호, 김경하, and 이명훈, "스마트 축사환경 데이터 기반 가축 최적 생육 환경 조성 시스템에 관한 연구," 한국통신학회 학술대회 논문집, pp. 402-403, 제주, 2022-06-22.
- [5] 강성원, 펜한 안토니오 엔 림 (임대한), and 박정희, "연합 학습 환경에서 이상치 탐지를 위한 Isolation Forest 학습," 한국정보과학회 학술발표논문집, pp. 985-987, 제주, 2022-12-20.