

품질관리 기준 정립을 위한 양돈 데이터 전처리 기법 연구

이지은, 이명훈*

*국립순천대학교

eunicejlee03@gmail.com, *leemh777@scnu.ac.kr

Data Preprocessing Techniques for Swine Production: Toward the Establishment of Quality Control Standards

Lee Ji Eun, Lee Meong Hun*

*Sunchon National Univ.

요약

본 연구는 양돈 농장의 환경 센서 데이터를 활용하여 데이터 품질관리 기준 수립을 위한 전처리 절차를 제안한다. 연구 대상은 온도, 상대습도, 이산화탄소, 암모니아 등 환경 변수이며, 실제 농장 데이터의 특성을 반영해 결측치와 스파이크형 이상치를 인위적으로 주입하였다. 전처리 과정은 결측치 보완, 이상치 교정, 정규화의 세 단계로 구성되며, 선형 보간과 중앙값 대체, Hampel 필터(중앙값 $\pm k \cdot 1.4826 \times \text{MAD}$), Robust Z-score 등의 기법을 적용하였다. 평가 결과, Hampel 필터와 보간 결합 방식이 이상치 교정에서 가장 낮은 RMSE를 보였으며, 정규화 이후 변수 간 스케일 불일치가 해소되어 분석 신뢰도가 향상되었다. 본 연구는 양돈 환경 데이터의 품질 문제 유형을 체계화하고, 실환경 오류를 모사한 정량적 검증을 통해 전처리 기준의 실효성을 검증하였다는 점에서 의의가 있다. 나아가 제시된 결과는 축산 데이터의 실시간 품질관리 및 표준화 체계 구축의 기초자료로 활용될 수 있다.

I. 서론

현대 양돈 산업에서는 ICT 기반 환경센서를 통해 온도, 습도, 이산화탄소(CO_2), 암모니아(NH_3) 등의 주요 지표가 실시간으로 수집된다. 이러한 데이터는 사육 환경의 정량적 제어, 질병 예측, 사양 최적화 등 다양한 응용에 활용되지만[1], 센서 오작동, 통신 지연, 수기 입력 오류로 인해 결측 및 이상치가 빈번히 발생한다[2-5]. 이러한 품질 저하는 지표 간 비교를 어렵게 하고, 데이터 기반 의사결정의 신뢰도를 저하한다.

본 연구는 양돈 농장의 환경 센서 데이터를 대상으로 결측치 보완, 이상치 탐지·교정, 변수 정규화로 구성된 전처리 절차를 제시하고, 각 기법의 정량적 성능을 비교·평가함으로써 품질관리 기준 정립을 위한 기초 자료를 마련하고자 한다. 이를 위해 실환경에서 관찰되는 오류 양상을 모사한 평가용 데이터셋을 구축하고, Python 기반 자동화 파이프라인을 통해 동일 조건에서 전처리 기법을 적용하였다.

본 연구의 목적은 양돈 환경 데이터의 품질 문제를 정형화하고, 전처리 기법의 효과를 실증적으로 검증함으로써 향후 품질지표(QI) 정의 및 품질 관리 기준(QC guideline) 수립을 위한 기초 근거를 제공하는 데 있다.

II. 연구 방법

II.1. 데이터 수집 및 오류 주입

본 연구는 국내 양돈 농장의 이유자돈(weaning pig) 사육실에서 수집한 환경 데이터를 사용하였다. 온도, 상대습도, 이산화탄소(CO_2), 암모니아(NH_3) 센서를 네 개 챔버에 설치하고, 2023년 8월 8일부터 10월 4일까지 5분 간격으로 58일간 측정하였다.

원시 데이터의 품질이 양호해 실제 결측 및 이상치가 거의 없었으므로, 평가용 오류를 인위적으로 주입하였다. 결측은 과일당 약 20포인트의 연속 구간으로 삽입하였고, 이상치는 온도 컬럼별 12포인트의 스파이크형 값을

추가하였다. 이상치 위치는 중앙값 기준 편차가 큰 구간을 우선 선택하였으며, 편차 계수 k 를 3~6 범위로 설정해 민감도를 조정하였다.

또한 센서별 결측률, 표준편차, 강건 정규화 오차를 산출하여 신뢰성을 비교하였고, 그 결과 T_1.5 센서를 기준 신호로 활용하였다.

II.2. 전처리 파이프라인

전처리 과정은 결측치 보완, 이상치 교정, 정규화의 세 단계로 구성하였다. 결측치 보완은 시계열의 연속성을 유지하고 추세 왜곡을 최소화하기 위해 유형별로 기법을 구분하였다. 짧은 구간의 연속 결측(≤ 5 포인트)은 인접 구간의 기울기를 반영한 선형 보간으로, 산발 결측은 변수별 중앙값으로 대체하였다. 이 조합은 데이터의 구조를 보존하면서 계산 효율을 높일 수 있다.

이상치 교정은 물리적 범위 검증 후 Hampel 필터(중앙값 $\pm k \cdot 1.4826 \times \text{MAD}$)와 변화율 규칙을 결합해 스파이크형 이상치를 탐지하였다. 검출된 이상치는 선형 보간, Hampel 대체, Clip 세 가지 방식으로 교정하였으며, 급격한 변동을 완화하면서 원 신호의 위상과 추세를 유지하도록 하였다.

정규화는 변수 간 단위와 범위 차이를 제거하기 위해 Robust Z-score를 적용하였다. 중앙값과 $1.4826 \times \text{MAD}$ 를 이용해 극단값 영향을 완화하고, 모든 변수를 평균 0, 분산 1의 공통 스케일로 변환하였다. 이를 통해 다변량 분석과 품질 지표 산출 시 변수 간 비교 가능성을 확보하였다.

II.3. 평가 지표 및 분석 방법

전처리 성능 평가는 오류 주입 구간의 원래 값과 복원값 간의 RMSE를 기준으로 수행하였다. 결측치 보완 및 이상치 교정 각각에 대해 동일한 구간에서 RMSE를 산출하고, 평균 오차의 크기를 비교함으로써 전처리 정확도를 정량적으로 평가하였다.

정성적 평가는 시각화 분석을 병행하여 수행하였다. 전처리 전후의 시계열 데이터를 동일 축상에 오버레이하여 추세 및 위상 보존 여부를 확인하였

고, 히스토그램과 커널 밀도 추정(KDE)을 이용해 분포 정렬 효과를 분석하였다. 또한, 중앙값과 산포도의 변화를 비교하여 전처리 이후 데이터의 구조적 안정성과 스케일 통일 여부를 검증하였다.

이러한 정량·정성 평가 결과를 종합하여 전처리 기법 간 성능 차이를 분석하고, 향후 품질지표 설계 시 활용 가능한 기준 지표로 RMSE와 분포 중심 편차 등을 제안하였다.

III. 분석 결과

III.1 결측치 보완 결과

선형 보간은 중앙값 대체 대비 RMSE가 약 40~60% 낮았으며, 결측 구간 주변의 기울기와 위상을 연속적으로 복원하였다. 이는 시계열의 지역적 선형성 가정이 깊은 연속 결측에서 유효하게 작동하기 때문이며, 분포 중심만을 이용하는 중앙값 대체와 달리 변동 패턴의 보존에 유리하다. 그림 1은 결측치 대체 기법별 RMSE 비교 결과를 보여주며, 선형 보간이 깊은 결측 구간에서 가장 낮은 오차를 나타낸을 확인할 수 있다. 결측 구간이 길어질수록 오차가 증가하는 경향이 나타나, 스플라인 또는 로컬 회귀 기반의 고차 보간법이 보완 대안으로 고려될 수 있다.

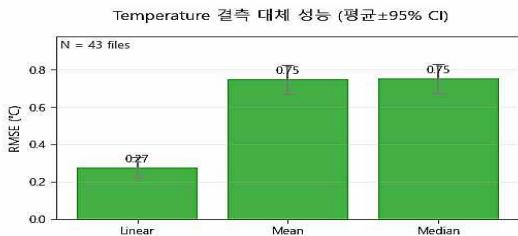


그림 1. 결측치 대체 성능 비교

Fig. 1. Comparison of missing value imputation performance

III.2 이상치 교정 결과

스파이크형 이상치 교정에서는 Hampel 필터와 선형 보간의 결합이 가장 우수하였다. Hampel 필터는 이동 창 내 중앙값과 절대편차(MAD)를 이용해 극단값을 안정적으로 식별하고, 이후 인접 샘플을 이용한 선형 보간이 신호의 위상·진폭을 원래 수준으로 복원하였다. 그림 2는 이상치 주입 및 교정 전후의 시계열을 비교한 것으로, Hampel 필터와 보간 결합 방식이 급격한 변동을 효과적으로 완화하고 원래의 추세를 유지함을 시각적으로 보여준다. 이 방식은 클리핑이나 평균 대체보다 약 25% 낮은 RMSE를 보였으며, 변화율 기반 규칙을 병행할 경우 단기 급변의 오검출을 줄여 탐지의 특이도를 높이는 것으로 나타났다.

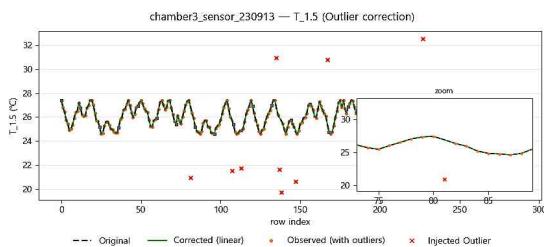


그림 2. 이상치 주입 및 교정 결과 시각화

Fig. 2. Visualization of outlier injection and correction results

III.3 정규화 효과 및 품질 논의

정규화 수행 후 변수들의 분포 중심은 0, 분산은 1에 수렴하여 센서 간 단위와 범위의 불일치가 해소되었다. 그림 3은 정규화 전후의 데이터 분포 변화를 나타내며, Robust Z-score 적용 후 극단값의 영향이 감소하고 각 센서 변수가 공통 스케일로 정렬됨을 확인할 수 있다. 이 과정은 다변량 분석의 안정성과 비교 가능성을 향상시켰으며, 전처리 단계별 정량지표와 분포·

오버레이 시각화를 종합하여 산출한 품질지표는 향후 품질관리 임계값 설정과 자동 점검 규칙 설계의 기준값으로 활용 가능성이 있다.

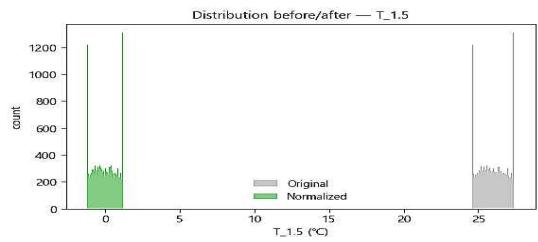


그림 3. 정규화 전후 분포 비교

Fig. 3. Comparison of data distribution before and after normalization

IV. 결론

본 연구에서는 양돈 사육실 환경 데이터를 대상으로 결측, 이상치, 정규화의 전처리 절차를 제안하였으며, 각 단계의 성능을 정량·정성적으로 평가하였다.

연구 결과, 선형 보간은 깊은 결측 구간의 복원 정확도가 높아 기본 전처리 기법으로 적합하고, Hampel 필터와 보간 결합 방식은 스파이크형 이상치를 효과적으로 제거하면서 신호의 추세를 안정적으로 유지하였다. 또한, Robust Z-score 정규화를 통해 변수 간 스케일 불일치가 해소되어 데이터의 일관성과 해석 가능성이 향상되었다.

이러한 결과는 전처리 기법의 상대적 성능을 실증적으로 비교한 것으로, 향후 품질관리 기준 정립과 품질지표 정의를 위한 기초 자료로 활용될 수 있다. 향후 연구에서는 계절성과 주기성을 반영한 확장형 전처리 모델을 개발하고, 실시간 품질관리 체계를 연계하여 데이터 품질 관리의 자동화 기반을 마련하고자 한다.

ACKNOWLEDGMENT

이 논문은 2025년도 전라남도 재원으로 전남인재평생교육진흥원 지원을 받아 수행됨.

참 고 문 헌

- [1] 김민지, 모창연, 홍순종, 김현태, 조병관, 이대현, 신창섭, 장경제, “ICT 기술을 접목한 양우 스마트 축산 연구 동향 및 기술 동향 분석,” 한국 농업기계학회 학술발표논문집, vol. 26, no. 1, pp. 167, 2021년 4월.
- [2] 고경일, 이명훈, “축사 환경 데이터 신뢰성 확보를 위한 이상 탐지 알고리즘 적용,” 2024년도 한국통신학회 추계종합학술발표회 논문집, pp. 379-380, 2024년 11월.
- [3] 문준섭, 박정환, 김락우, 이승현, 김찬민, 석희웅, 안수빈, 이선향, 김한, “돈사 내부 가상센서 적용을 위한 기계학습 모델 설계 및 분석,” 한국 농공학회 학술대회초록집, 2023권, pp. 232 - 232, 2023년 10월.
- [4] 정광훈, 여현, “Isolation Forest를 활용한 축산 온·습도 이상치 탐지에 관한 연구,” 한국통신학회, 2024년도 한국통신학회 하계종합학술발표회 논문집, pp. 296 - 298, 2024년 1월.
- [5] 김승강, 남기표, “스마트팜 구축을 위한 데이터수집의 문제점 분석 연구,” 융합보안논문지 (Convergence Security Journal), vol. 22, no. 5, pp. 69 - 80, 2022년.