

상수도관 누수 탐지 신뢰성 향상을 위한 누수 파라미터 손실 대처 기술

홍영기°, 채지영°, 박경준°°°

대구경북과학기술원 인공지능학과°, 전기전자컴퓨터공학과°, 전기전자컴퓨터공학과°°°

hykfun@dgist.ac.kr°, cowldud3@dgist.ac.kr°, kjp@dgist.ac.kr°°°

1. 서론

일상생활에서 담수 부족으로 인해 다양한 지하 파이프라인 네트워크에서 누수 감지는 매우 중요하다. Internet of Things(IoT)을 이용하여 누수를 정확하게 감지하는 것은 필수적인 연구로 인식된다.[1] 하지만 IoT의 네트워크 및 하드웨어 문제로 인해 IoT가 누수 파라미터를 전송 또는 저장을 못 하는 상황이 발생할 수 있다. 이러한 누수 파라미터의 손실은 누수 진단 모델의 성능에 지장을 줄 수 있다. 따라서 높은 누수 진단 모델의 정밀도를 보장하기 위해서는 불완전한 누수 파라미터들을 추정하여 완전한 누수 파라미터로 구축할 수 있는 누수 파라미터 추정 모델 구축이 필요하다.

2. 본론

본 연구에서는 누수 파라미터 손실에 대해서 누수 파라미터 대처 기술을 이용한 효율적인 누수 탐지 시스템 개발을 위한 비교분석을 했다. 제안하는 시스템은 전체 누수 파라미터의 10%, 30%, 50% 비율로 누수 파라미터 무작위 손실을 발생시켰으며, 다양한 시나리오의 데이터에 적용 가능하다. 무작위 손실은 일시적인 센서 중단으로 인해 나타나는 흔한 결측 데이터의 패턴이다 [2]. 제안하는 시스템의 성능을 비교하기 위해 완전한 데이터를 이용하여 Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machine (SVM), Multilayer Perceptron (MLP) 을 사전에 수행했다. 데이터 샘플의 불균형 문제로 인해 샘플 수의 비중에 따라 정밀도와 재현율의 가중평균을 한 Weighted F1-Score 성능 지표를 이용했다. 각 모델의 Weighted F1-Score 는 83.1%, 77.5%, 78.4%, 76.5%, 80.1% 를 가진다. 누수 파라미터 추정 기법은 Mean Imputation, KNN Imputation, Auto Encoder (AE) Imputation 을 사용했다. 누수 파라미터 추정 기법의 성능은 Root Mean Squared Error (RMSE)를 사용하여 평가했다. 이후 위 3 가지 누수 파라미터 추정 기법을 적용한 파라미터 셋을 이용하여 누수 탐지 성능을 평가했다. 연속형의 파라미터 셋에서 Mean imputation 기법 같은 경우, 가장 흔한 방법으로 각 손실한 값을 해당 변수의 평균으로 대체하는 것이다. KNN Imputation 기법 같은 경우, 하나의 결측 데이터를 중심으로 가장 가까운 K 개의 데이터를 선택하여 이것들의 평균치로 대체하는 기법이다. 이때 유클리드 거리를 사용하여 가장 가까운 K 개의 파라미터를 선택한다. AE Imputation 인 경우, Encoding

함수를 통해 저차원으로 Mapping 다음 저차원 공간에서 다시 Decoding 함수로 파라미터 값을 추정하는 방식이다. 누수 파라미터 추정 3 가지 기법을 토대로 5 가지 분류 모델을 사용하여 시나리오를 수행했다. 먼저 누수 파라미터 추정 3 가지 기법의 성능을 보자면, 3 가지 기법 중에 KNN Imputation 기법의 RMSE 값은 누수 파라미터 비율에 따라 각각 0.013, 0.024, 0.035 를 가진다. 반대로 Mean Imputation 기법의 RMSE 값은 0.108, 0.368, 0.788 을 가진다. 특히 Missing value 의 50%에서는 상대적으로 매우 높은 값을 보여준다.

다음으로 위 3 가지 파라미터 추정 기법을 사용한 데이터 세트를 이용하여 5 가지 분류 모델의 성능을 살펴보면, KNN Imputation 추정 기법 기반 분류 모델 중 DT 인 경우 손실된 데이터 비율이 증가할수록 DT 분류 모델의 Weighted F1-Score 의 감소가 있지만, 다른 두 기법에 비해 상대적으로 성능을 잘 유지한다. KNN Imputation 추정 기법 기반 DT 모델의 Weighted F1-Score 는 81.3%, 79.8%, 78.5%를 가진다. Mean Imputation 추정 기법은 데이터 누락 비율이 증가함에 따라 성능이 저하되는 경향이 비교적 뚜렷하다. Mean Imputation 추정 기반 DT 모델의 Weighted F1-Score 는 77.1%, 76.4%, 72.1%를 가진다. Mean Imputation 기법은 데이터의 전반적인 경향만을 반영하기 때문에, 누수 파라미터의 복잡성이나 특정 패턴을 충분히 포착하지 못하는 단점이 드러난다.

3. 결론

본 연구는 누수 파라미터의 손실이 발생할 때 적절히 대응할 수 있는 데이터 대체 기술을 적용한 효율적인 누수 탐지 시스템을 위한 비교 분석을 했다. 평균적으로 KNN Imputation 기반 분류 모델 중 DT가 80.1%로 가장 높은 성능을 가진다. 이는 IoT 장치가 예상대로 작동하지 않거나 데이터 손실이 발생하는 상황에서도 신뢰성 있는 누수 탐지 모델을 구축할 수 있는 새로운 가능성을 제시한다. 기술적 장애나 예기치 못한 데이터 손실 상황에서도 수자원 관리의 연속성과 신뢰도를 유지할 수 있는 강력한 도구의 역할을 할 것으로 기대한다.

4. 참고 문헌

[1] Jirhandeh, Mohammad Askari. *Leak Detection of Water Pipeline Networks with Acoustic Dataset Analysis*. MS thesis. 2021.

[2] Ehrlinger, Lisa, et al. "Treating missing data in industrial data analytics." *2018 Thirteenth International Conference on Digital Information Management (ICDIM)*. IEEE, 2018.