

트리 기반 부스팅 알고리즘을 이용한 상수도관 누수 탐지 방법

이재흥^{1*}, 오윤성², 민준혁²

¹한밭대학교 컴퓨터공학과 교수, ²한밭대학교 컴퓨터공학과 학생

Leakage Detection Method in Water Pipe using Tree-based Boosting Algorithm

Jae-Heung Lee^{1*}, Yunsung Oh², Junhyeok Min²

¹Professor, Department of Computer Engineering, Hanbat National University

²Student, Department of Computer Engineering, Hanbat National University

요약 국내 상수도관의 파열, 결함 등으로 인한 누수율로 인한 손실이 매우 크고, 이런 누수를 예방을 위한 방지 대책이 필요한 상황이다. 본 논문에서는 진동 센서를 활용한 누수 탐지 센서를 개발하고 인공지능 기술을 활용한 최적의 누수 탐지 알고리즘을 제시하고자 한다. 상수도 배관에서 취득한 진동음은 FFT(Fast Fourier Transform)를 이용한 전처리 과정을 거친 뒤, 최적화된 트리 기반 부스팅 알고리즘을 적용하여 누수 분류를 하였다. 다양한 실증 환경에서 취득한 약 26만여 개의 실험 데이터에 적용한 결과 기존의 SVM(Support Vector Machine) 방법에 비해 약 4%가 향상된 97%의 정확도를 얻었고, 연산 처리속도는 약 1,362배가 향상되어 엣지 디바이스 적용에도 적합함을 확인하였다.

주제어 : 누수 탐지, 상수도관, FFT(Fast Fourier Transform), 트리 기반 알고리즘, 주파수 스펙트럼

Abstract Losses in domestic water supply due to leaks are very large, such as fractures and defects in pipelines. Therefore, preventive measures to prevent water leakage are necessary. We propose the development of a leakage detection sensor utilizing vibration sensors and present an optimal leakage detection algorithm leveraging artificial intelligence. Vibrational sound data acquired from water pipelines undergo a preprocessing stage using FFT (Fast Fourier Transform), followed by leakage classification using an optimized tree-based boosting algorithm. Applying this method to approximately 260,000 experimental data points from various real-world scenarios resulted in a 97% accuracy, a 4% improvement over existing SVM(Support Vector Machine) methods. The processing speed also increased approximately 80 times, confirming its suitability for edge device applications.

Key Words : Leakage detection, Water pipes, FFT(Fast Fourier Transform), Tree-based algorithms, Frequency spectrum

1. 서론

상수도관은 양질의 물을 안전하게 공급하는 것을 목표

로 한다. 이를 위해 적절한 수압과 유량이 유지되어야 한다. 하지만 예측 불가능한 파열과 결함 등으로 인해서 설
계된 기능을 모두 발휘하지 못하고 있다. 2011년 정수장

본 논문은 산업통상자원부 및 한국산업기술기획평가원(KEIT) 연구비 지원에 의한 연구임.

*교신저자 : 이재흥(jhlee@hanbat.ac.kr)

접수일 2024년 03월 04일 수정일 2024 03월 18일 심사완료일 2024년 03월 29일

에서 생산된 6,021 백만 m^3 의 수도물 중 요금이 징수된 유수 유량은 약 4,920 백만 m^3 정도이다. (유수율 83.2%) 선진국에 비해 매우 낮은 수준이다[1]. 하지만 적절하게 상수도관의 상태를 감지할 누수 탐지 센서와 센서로부터의 데이터를 분석할 누수 탐지 알고리즘이 부재한 상황이다.

따라서 본 논문에서는 다양한 재질의 상수도관에서 누수 센서를 통해 취득한 진동음 데이터를 FFT(Fast Fourier Transform)를 거쳐 주파수 스펙트럼으로 변환하고, 이를 배관음의 특징에 따라 누수음과 일반음으로 분류하는데 필요한 최적화된 트리 기반의 인공지능 알고리즘을 제시하고자 한다. 또한, 최근 확산되고 있는 원격 스마트 미터링 기능과 함께 적용 가능하도록 경량화된 모델을 제시함으로써 온 디바이스 AI 누수탐지기에 적합하도록 한다.

2. 기존의 누수 탐지 방법

기존 진행된 연구 사례 몇 가지를 살펴보면 우선, 수자원공사의 관망 감시 CCTV 시스템을 활용하여 누수 사고 발생을 검출하기 위한 GAN(Generative Adversarial Networks) 기반의 감지 시스템 구현 시도가 있었지만 외부로 노출된 누수만 감지할 수 있는 한계가 있었다[2]. 유량 계측기를 활용한 시계열 데이터 기반에 순환신경망(RNN, Recurrent Neural Network)의 Vanishing Gradient 문제를 해결한 LSTM(Long-Short Term Memory networks) 알고리즘을 기반으로 한 누수 감지 시도가 있었다[3]. 하지만 LSTM은 많은 연산이 필요하고, 유량을 통해서 누수를 판단하기 때문에 물 수요가 많으면 누수의 인지가 부정확해지는 한계가 있었다. 배관에서 유량을 Neuro-Fuzzy 알고리즘을 활용하여 누수 예측을 한 시도도 있다[4]. 이 또한 유량을 이용했을 경우의 한계점을 공유한다. 배관 수압 변화를 Bayesian method를 활용한 통계적 시스템을 시도하거나[5], XGBoost와 같은 알고리즘을 사용하여 압력 변화에 따른 누수 여부 판단을 시도한 연구 또한 존재한다[6]. 이 경우에도 물 수요에 민감하다는 단점이 있다. 파이프에서 수집된 음향신호를 이용한 연구로는 센서를 통해 수집된 음향신호에 대한 MFCC(Mel-Frequency Cepstral Coefficient)를 적용해 획득한 이미지 데이터 기반에 CNN(Convolutional Neural Network) 알고리즘을 활용한 모델을 사용하는 시도가 있었다[7]. 하지만 CNN

알고리즘은 연산량이 많다는 단점이 있다. 또한 DT(Decision Tree)에 Ensemble 기법을 적용한 LightGBM, XGBoost, CatBoost, AdaBoost, Random Forest 등의 알고리즘을 비교 실험한 사례가 있으며[8] GBT(Gradient Boosting Tree)를 이용한 MEL(Multi-Strategy Ensemble Learning) 전략을 활용하는 시도 또한 있었다[10]. 마지막으로 배관의 진동음에 대한 스펙트럼을 데이터로 하여 나타나는 특징을 SVM(Support Vector Machine)을 이용해 분류하려는 시도가 있었다[9]. SVM은 군집이 존재하는 데이터셋에서 임의의 데이터를 군집으로 분류하는 것을 목표로 하는 알고리즘이며, 분류를 위해서 임의의 데이터를 m 차원의 점으로 간주한 후 데이터셋의 샘플을 고차원인 M 차원에 대응시킨 뒤 군집을 분할하는 초평면 경계를 만든다. 이 과정에서 군집 사이의 최소 거리를 최대로 만드는 초평면을 선택하게 되는데 그 결과로 새로운 데이터에 강건한 장점을 가지나 데이터가 고차원이 되므로 연산량이 많아져 옛지 디바이스에는 적절하지 못하다.

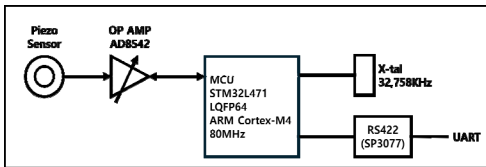
트리 기반 알고리즘은 임의의 변수 영역을 두 개로 구분하는 재귀적 분기(recursive partitioning)와 너무 많은 분할을 억제하는 가지치기(pruning) 과정을 통해 학습을 진행하며, 추론 시에는 분기에 따른 분류만 진행하므로 연산량이 적다. 대표적 예시로는 결정 나무(Decision Tree)[11], LightGBM[14], XGBoost[15] 등이 있다. 이 중 결정 나무 알고리즘은 단일 트리 구조이기 때문에 특정 데이터에만 잘 작동할 수 있는 한계가 있다. 따라서 여러 학습기를 결합해 성능을 개선하는 앙상블(Ensemble)[12] 기법과 이전 모델에서 발생한 오차를 중심으로 학습을 진행해 높은 예측 성능을 달성하는 GBDT(Gradient Boosting Decision Tree)[13]가 제안되었다. 본 논문에서는 배관의 진동음에 대한 스펙트럼 데이터를 사용하여, 적은 연산 능력과 전력 소모 같은 특징을 가지는 옛지 컴퓨터의 특성에 더욱 부합하는 알고리즘으로 CNN 기반[7] 또는 SVM 기반[9] 누수 탐지 대신 트리 기반의 알고리즘을 활용해 이전보다 더 빠른 분류 속도를 가지며 더 높은 분류 성능을 보이는 모델을 제시한다.

3. 본 논문의 누수 탐지 방법

3.1 누수 탐지 센서

본 논문에서는 상수도관의 진동 신호에 대한 주파수

스펙트럼을 추출하기 위해 수도관에 장착 가능한 누수 탐지 센서를 개발하였다. 누수 탐지 센서는 [Fig. 1]과 같이 수도관에서 발생하는 진동을 압전 센서를 통해 얻은 미세한 진동 신호를 증폭과 노이즈 제거 필터를 거쳐 ADC(Analog-to-Digital Converter)를 통해 디지털 신호 레벨로 변환한다. 본 논문에서는 약 1428.57Hz의 주기로 2000번 샘플링 하여 진행하였다. 획득된 시간-파형 데이터를 FFT(Fast Fourier Transform)를 진행해 1Hz 간격으로 0Hz~499Hz까지 총 500개의 주파수 대역의 데이터를 생성 후 RS422 프로토콜을 이용하여 클라이언트에 변환된 스펙트럼 데이터를 전송한다[Fig 2].



[Fig. 1] Sensor module diagram

Algorithm 1: Sensor module operation

```

init state: sampling repeat = 2000
            raw_ADCdata = [];
1 for i = 1 to sampling repeat do
2   tmp = get_pizo_sinal();
3   Digital_signal = Analog-to-digital(tmp);
4   raw_ADCdata.append(Digital_signal);
5 end
6 frequency elements = FFT(raw_ADCdata);
7 return frequency elements;
    
```

[Fig. 2] Sensor module operation

3.2 트리 기반의 누수 탐지 알고리즘

본 논문에서는 주파수 스펙트럼 데이터를 트리 기반 기계학습 알고리즘 중 LightGBM[14], XGBoost[15]라는 알고리즘을 사용한다. 트리 기반 알고리즘은 데이터 분류 시 분기점에 대한 비교 연산만 수행하여 부동소수점 연산이 많은 SVM과 같은 알고리즘 대비 분류 속도가 매우 빠르며, 모든 데이터를 고려하여 최적의 트리 분할 지점을 찾는 기존의 GBDT 알고리즘보다 적은 연산으로 적절한 분할 지점을 탐색하기 때문에 연산 장치의 제약이 존재할 수 있는 옛지 컴퓨터에 적합하다.

3.2.1 LightGBM

LightGBM은 기존의 GBDT 알고리즘을 개선하고자 GOSS(Gradient-based One-Side Sampling)와 EFB(Exclusive Feature Bundling)를 제안하였다. GOSS

와 조합하면 $O(\text{데이터 부분집합 수} \times \text{특성 수})$ 로 많은 계산량을 절약할 수 있다.

3.2.1.1 GOSS (Gradient-based One-Side Sampling)

GOSS는 입력 데이터 중 기울기가 작은 것은 이미 모델이 잘 학습한 것으로 간주하고, 기울기가 높은 데이터를 추출해 학습에 활용하는 방법이다. 이때 기울기가 큰 데이터로 인해서 모델이 불안정해지는 것을 막기 위해서 $\frac{1-a}{b}$ 를 무작위 샘플에 곱하여 값을 보정한다. 이때 a 는 기울기가 큰 데이터의 샘플링 비율, b 는 기울기가 작은 데이터의 샘플링 비율이다. 전체 데이터의 부분집합을 이용하므로 연산량이 크게 줄어드는 효과가 있다.

3.2.1.2 EFB (Exclusive Feature Bundling)

EFB는 입력 데이터의 특징 수를 감소시키는 전략이다. 상호 배반적인 특성들을 번들링 하며, 번들링 하고자 하는 특징이 공통적인 범위를 가진다면, 한쪽에 오프셋을 더하여 정보의 손실을 방지한다. 해당 방법을 이용하여 본 논문에서 누수 탐지에 사용되는 500개 특징을 가진 고차원 음향 데이터를 저차원 데이터로 변환한다면 시간 복잡도가 줄어드는 효과가 있다.

3.2.2 XGBoost

XGBoost는 GBDT 알고리즘을 개선하기 위해 과적합을 방지하며 이전 학습 결과를 기반으로 가중치를 부여하고, 최적의 분할 지점을 탐색을 제안하여 강력한 모델을 빠르게 구축할 수 있는 알고리즘이다. 또한 트리를 균형 분할하기 때문에 학습 데이터가 증가할 때 트리가 직렬화되고 모델이 커지는 문제가 발생하지 않아 정확도가 우수하고 경량 컴퓨팅 기기를 이용한 누수 탐지에 적합하다.

3.2.2.1 Regularized Loss Function

Regularized Loss Function은 이전 학습 결과를 기반으로 학습이 부족한 부분에 가중치를 부여하고 리프 노드의 수를 줄여 과적합을 방지한다. 해당 함수는 다음과 같다.

$$L(\varnothing) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \quad \text{where } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

[Fig. 3] Regularized Loss Function

여기서 T 는 f_k 의 리프 수, w 는 리프의 값, l 은 실

제 값과 예측값의 차이를 계산하는 미분 가능한 Convex 손실 함수이다.

3.2.2.2 최적 분할 지점 탐색

분할 지점 탐색 시 가능한 모든 지점에 대해서 손실 값 차이를 계산한다면 가장 정확도가 높은 트리를 얻을 수 있지만 계산량이 많고 메모리 사용이 많다. XGBoost는 대안으로 데이터를 분할 지점 후보로 나뉜 빈(bin)에 넣은 뒤 분할 지점을 계산하여 최적 분할 지점의 근사치를 구하여 상대적으로 빠른 속도로 많은 데이터를 학습시킬 수 있다.

4. 실험 및 결과

4.1 데이터 수집 환경

본 논문에서는 누수 데이터 측정을 위해 다음의 2가지 환경에서 취득하였다. 첫 번째 환경은 외부 환경에 의해 데이터에 잡음이 생기는 것을 줄이기 위해 실내에 조성되었으며, 누수 환경을 조성하기 위해 파이프의 다양한 지점에 구멍을 뚫어 누수 지점을 만든 후 밸브를 이용해 누수 여부를 조작한다. 본 실험에서 구축한 테스트베드의 주요 제원은 <Table 1>과 같으며 수집된 데이터는 누수음 34,601개, 정상음 157,879개다.

<Table 1> Indoor simulation environment specifications

Category	Unit	Value	
Pipe type	-	PB	Steel
Inside diameter	mm	15	
Pipe length	m	1.6	2
Pipe shape	-	Curved	Straight
Diameter of leak point	mm	1	4
Number of leak points	-	3	1
Distance between leak points	cm	24.5	-

두 번째 환경은 대전광역시 유성구에 위치한 K-water 인재개발원 내부의 상수도관 모형 플랜트이다. 본 시설에서도 마찬가지로 누수 환경을 시뮬레이션하기 위해 각 파이프에 구멍을 뚫어 누수 지점을 생성한 후 밸브를 이용해 누수 여부를 조작한다. 본 실험에서 사용한 테스트베드의 주요 제원은 <Table 2>와 같으며, 수집된 데이터는 누수음 535개, 정상음 7,889개다.

<Table 2> K-water simulation environment specifications

Category	Unit	Value		
		DCIP	SP	PE
Pipe type	-	DCIP	SP	PE
Pipe diameter	mm	100		
Pipe length	m	300		
Flow rate	m/s	1		
Diameter of leak point	mm	1, 2, 3, 4		
Number of leak points	-	1, 1, 1, 1		
Distance between leak points	cm	60-70		

위 환경의 각 배관에 누수 센서를 설치 후, 누수 센서의 1회 측정 시마다 감지된 진동에 대한 주파수 대역의 데이터를 어레이 형태로 PC에 전송받고, 이를 500개 행의 CSV(Comma-Separated Values) 파일로 저장하는 과정을 반복하여 데이터를 저장한다. 측정된 데이터는 밸브 개폐 여부에 따라 정상(NORMAL)과 누수(LEAK) 2가지 클래스로 구분하였다.

마지막으로 AI-Hub(<https://aihub.or.kr/>)에서 정부 공공 데이터센터에서 제공하는 상수도관 누수 감지 데이터를 앞서 측정된 데이터와 함께 사용하였다. 본 데이터는 광주광역시 및 전라남도 고흥군 등 일부 지역에서 측정되었다. 수도 계량기함 및 밸브실에 센서를 설치 후 수집된 데이터를 현장 검증을 통해 품질관리를 진행하였다. 그렇게 검증을 마친 데이터는 옥내누수(In), 옥외누수(Out), 기계·전기음(Noise), 환경음(Other), 정상(Normal)의 5가지 클래스로 구분하고 있다<Table 3>. 또한 본 논문에서 사용된 누수 센서와 유사하게 상수도관에서 측정된 진동 데이터를 10Hz 간격으로 0Hz~ 5120Hz까지 총 513개의 주파수 대역으로 변환하여 제공한다.

<Table 3> AI-Hub Data

Class	Number of data		
	Train	Validation	Total
Out	17,540	2,193	19,733
In	13,274	1,660	14,934
Noise	5,030	630	5,660
Other	7,020	879	7,899
Normal	19,705	2,463	22,168
Total	62,569	7,825	70,394

본 실험에서는 AI-Hub에서 제공한 데이터를 이용한 선행연구와[8] 알고리즘 성능 비교를 위해 해당 데이터를 이용해 SVM과 트리 기반 모델 간의 성능을 비교한다. 그리고 실내 조성 환경 및 K-water에서 자체 수집한 데이터로 추가 비교를 실시하였다.

4.2 데이터 전처리

누수 센서에서 수집된 데이터와 AI Hub 공개데이터의 주파수 간격과 대역폭이 다르기 때문에, 두 유형의 데이터에서 공통으로 보유하고 있는 주파수 데이터만 실험에 사용한다. 또한 두 데이터 유형 모두 0Hz의 데이터는 사용되지 않는다. 즉, 알고리즘의 학습 및 검증에서는 10Hz 간격으로 10Hz~490Hz까지 49개의 주파수 데이터를 사용한다. AI Hub 데이터의 옥내누수(IN)와 옥외누수(OUT)를 제외하고 나머지 3개의 클래스는 정상(NORMAL) 상태로 나타내고, 누수 센서로 수집된 데이터는 2개의 클래스로만 데이터를 분류하므로, 본 실험에서는 AI Hub 데이터의 정상(NORMAL) 클래스를 정상음 데이터로, 옥외누수(OUT)와 옥내누수(IN) 클래스를 누수(LEAK)로 재 라벨링 해 활용한다.

본 실험에서는 주어진 데이터를 7:3의 비율로 학습 데이터와 검증 데이터로 분할하였다. 정보 불균형을 막기 위해 분할 시 각 클래스가 균등한 비율이 되며 중복을 허용하지 않게 무작위 추출하여 학습 데이터와 검증 데이터를 아래와 같이 생성하였다.

〈Table 4〉 Indoor Data

Class	Number of data		
	Train	Validation	Total
Leak	24,221	10,380	34,601
Normal	110,515	47,364	157,879
Total	134,736	57,744	192,480

〈Table 5〉 K-water Data

Class	Number of data		
	Train	Validation	Total
Leak	374	161	535
Normal	5,522	2,367	7,889
Total	5,896	2,528	8,424

〈Table 6〉 AI-Hub Data

Class	Number of data		
	Train	Validation	Total
Out	17,540	2,193	19,733
In	13,274	1,660	14,934
Noisse	5,030	630	5,660
Other	7,020	879	7,899
Normal	19,705	2,463	22,168
Total	62,569	7,825	70,394

4.3 성능 지표

본 실험에서 사용한 성능 지표는 다음과 같다. 정확도(accuracy)는 전체 샘플 중 옳게 예측한 샘플 수의 비율을 나타낸다. 정밀도(precision)는 예측한 클래스에서 실제 클래스와 일치하는 비율을 나타낸다. 재현율(recall)은 실제 클래스에서 예측한 클래스와 일치하는 비율을 나타낸다. F1 점수(F1-score)는 정밀도와 재현율의 조화평균이다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$recision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1score = \frac{2}{\frac{1}{recision} + \frac{1}{Recall}}$$

$$= 2 \times \frac{recision \times Recall}{recision + Recall}$$

[Fig. 4] Confusion Matrix

4.4 실험환경

본 논문에서 사용되는 엣지 컴퓨터는 Raspberry Pi 4 Model B 제품을 사용했으며, 학습 시 각 알고리즘의 하이퍼 파라미터 중 Random State는 11로 설정하였고, 결정 나무 알고리즘의 경우 깊이를 20으로 제한하였다. 나머지 하이퍼 파라미터는 모두 디폴트 값을 사용하여 학습 후 미리 분할된 테스트 데이터셋의 분류 정확도를 비교하였다. 알고리즘 및 성능 평가 관련 기능은 Python3의 패키지를 사용했다. 사용된 주요 패키지 및 기기 제원에 대한 정보는 〈Table 7〉과 같다.

〈Table 7〉 The specifications of edge computing device

Category	Description	
CPU	4 core 64-bit ARM-Cortex A72 1.5GHz	
GPU	VideoCore VI 3D Graphics	
RAM	LPDDR4 4GB	
OS	Linux 6.1.0-rpi6-rpi-v8, Debian GNU/Linux 12	
Version of libraries	Python	3.11.2
	Numpy	1.26.2
	Pandas	2.1.3
	Scipy	1.11.3
	Joblib	1.3.2
	Scikit-learn	1.3.2
	XGboost	2.0.2
LightGBM	4.1.0	

4.5 실험 결과

선행연구에서 사용된 알고리즘과 본 논문에서 제시한 알고리즘 간 성능 비교 결과는 다음과 같다.

<Table 8> The classification results of AI-Hub data

Category	Value			
	LightGBM	XGBoost	SVM	DT
Algorithm				
Training time (sec)	5.724	4.445	120.961	5.6
Inference time (sec)	0.431	0.057	44.64	0.015
Accuracy	0.95	0.96	0.92	0.93
Precision	0.95	0.96	0.93	0.93
Recall	0.95	0.96	0.92	0.93
F1-score	0.95	0.96	0.92	0.93

<Table 9> The classification results of Indoor testing data

Category	Value			
	LightGBM	XGBoost	SVM	DT
Algorithm				
Training time (sec)	17.537	20.346	967.598	56.435
Inference time (sec)	0.792	0.331	277.981	0.041
Accuracy	0.97	0.98	0.96	0.95
Precision	0.97	0.98	0.96	0.95
Recall	0.97	0.98	0.96	0.95
F1-score	0.97	0.98	0.95	0.95

<Table 10> The classification results of K-water testing data

Category	Value			
	LightGBM	XGBoost	SVM	DT
Algorithm				
Training time (sec)	3.249	3.27	1.221	1.581
Inference time (sec)	0.067	0.058	0.594	0.005
Accuracy	0.96	0.96	0.94	0.92
Precision	0.95	0.95	0.88	0.92
Recall	0.96	0.96	0.94	0.92
F1-score	0.95	0.95	0.91	0.92

<Table 11> The classification results of total data

Category	Value			
	LightGBM	XGBoost	SVM	DT
Algorithm				
Training time (sec)	20.586	20.479	2768.415	67.184
Inference time (sec)	1.78	0.187	1106.62	0.075
Accuracy	0.96	0.97	0.91	0.94
Precision	0.96	0.97	0.92	0.94
Recall	0.96	0.97	0.91	0.94
F1-score	0.96	0.97	0.91	0.94

실험 결과, 선행연구에서 사용된 AI-Hub 공개데이터 사용 시 트리 기반의 모델들이 최대 4% 높은 분류 성능을 보였으며, 학습 시간은 최대 27.21배, 추론 시간은 최대 2,976배 빠른 결과를 보였다. 자체적으로 수집한 데이터 중 K-water 데이터를 대상으로 DT가 SVM 대비 소폭 낮은 분류 성능을 보였으나, LightGBM과 XGBoost는 모든 데이터에 대해서 SVM 대비 높은 분류 성능을 달성하면서 학습 시간은 최대 135.18배, 추론 시간은 최대 5,917.74배 빠른 결과를 보였다.

5. 결론

상수도관에서 발생하는 진동음을 이용하여 누수를 판별하는 실험을 진행했다. 데이터는 파이프에서 발생하는 진동음을 압전 센서를 통해 측정 후 FFT를 통해 스펙트럼 데이터로 변환해 사용했으며, 실내 시뮬레이션 환경, K-water 모형 플랜트, AI-Hub 공개데이터를 활용했다. 옛지 기기에서 적용된 머신 러닝 모델을 분류 성능 및 소요 시간을 통해 평가하였다. LightGBM, XGBoost, SVM, DT 알고리즘을 적용했으며, 종합 결과 선행연구에서 사용된 SVM 대비 LightGBM, XGBoost가 정확도는 최대 약 4% 향상되어 최대 정확도 97%를 달성하였으며, 학습 시간 평균은 약 75배, 예측시간 평균은 약 1,362배 빠른 연산속도로 옛지 디바이스에 활용 가능한 것을 확인하였다.

REFERENCES

- [1] National Water Supply Information System, National water quantity management, https://www.waternow.go.kr/web/ssdoData/?pMENUID=8&ATTR_1=2011&ATTR_5=4
- [2] Seung-Heon Oh, Seung-hee Park, Ho-sung Kim, and Jong-rip Kim, "Implementation of GAN-based Water Pipeline Leakage Monitoring System," Korea Institute of information and Communication Engineering, Vol. 27, No. 1, pp. 32-35, 2023.
- [3] Lee, Chan Wook and Yoo, Do Guen, "Development of leakage detection model in water distribution networks applying LSTM-based deep learning algorithm," Journal of Korea Water Resources Association, Vol. 54, No. 8, pp. 599-606, 2021.
- [4] Jae-Moon Hwang, Ho-Hyun Lee, Gang-Wook Shin and Nam Kim. "Leakage Detection Prediction by Neuro-Fuzzy and WECC in Water Distribution Network.," Journal of Korean Institute of Intelligent

Systems, Vol. 27, No. 4, pp. 349-356, 2017.

- [5] Poulakis, Z., Dimitris Valougeorgis, and Costas Papadimitriou, "Leakage detection in water pipe networks using a Bayesian probabilistic framework." Probabilistic Engineering Mechanics Vol. 18, No. 4, pp. 315-327, 2003.
- [6] Xiaoqin Li, Xiaomei Wu, Mingzhuang Sun, Shengqiao Yang and Weikun Song, "A Novel Intelligent Leakage Monitoring-Warning System for Sustainable Rural Drinking Water Supply," Sustainability Vol. 14, No. 10, 2022.
- [7] Wei-Yi Chuang, Yao-Long Tsai, and Li-Hua Wang, "Leak Detection in Water Distribution Pipes Based on CNN with Mel Frequency Cepstral Coefficients," In Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence (ICIAI '19). Association for Computing Machinery, New York, NY, USA, pp. 83-86, 2019
- [8] Youngmin Seo, Kwanghyun Choi, Yuseong Lim, Byungjoon Lee and Yunyoung Choi "Application of Machine Learning Models for Water Pipeline Leakage Detection," Crisonomy, Vol. 19 No. 4, pp. 45-54, 2023.
- [9] Jungyu Choi and Sungbin Im, "Leak Detection and Classification of Water Pipeline based on SVM using Leakage Noise Magnitude Spectrum," Journal of the Institute of Electronics and Information Engineers, Vol. 60, No. 2, pp. 6-14, 2023.
- [10] Thambirajah Ravichandran, Keyhan Gavahi, Kumaraswamy Ponnambalam, Valentin Burtea and S. Jamshid Mousavi, "Ensemble-based machine learning approach for improved leak detection in water mains," Journal of Hydroinformatics, Vol. 23, No. 2, pp. 307-323, 2021.
- [11] Quinlan, J. R. "Induction of Decision Trees," Mach Learn, Vol. 1, pp. 81-106, 1985.
- [12] Tin Kam Ho, "Random decision forests," Proceedings of 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, Vol. 1, pp. 278-282, 1995.
- [13] Tianqi Chen, Hang Li, Qiang Yang and Yong Yu "General Functional Matrix Factorization Using Gradient Boosting," Proceedings of the 30th International Conference on Machine Learning, PMLR Vol. 28, No. 1, pp. 436-444, 2013.
- [14] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye and Tie-Yan Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc, Red Hook, NY, USA, pp. 3149-3157, 2017.
- [15] Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, pp. 785-794, 2016.

이 재 흥(Lee, Jae-Heung)

[정회원]



- 1985년 2월 : 한양대학교 전자공학과(공학석사)
- 1994년 2월 : 한양대학교 전자공학과(공학박사)
- 1989년 3월 ~ 현재 : 한밭대학교 컴퓨터공학과 교수

<관심분야>

임베디드 시스템, SoC 설계, 딥러닝

오 윤 성(Oh, Yun-Seong)

[준회원]



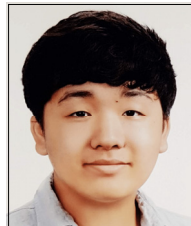
- 2024년 2월 : 한밭대학교 컴퓨터공학과(학사)

<관심분야>

임베디드 시스템, 인공지능

민 준 혁(Min, Jun-Hyeok)

[준회원]



- 2019년 3월 ~ 현재 : 한밭대학교 컴퓨터공학과

<관심분야>

임베디드 시스템, 인공지능