

Leak detection in water distribution pipes

using singular spectrum analysis

1. Introduction

- 물을 처리하고 전달하는 과정에서 많은 손실이 발생한다.
- 크게 leakage와 burst로 구분할 수 있다.
- 큰 burst는 발견하기 쉽지만, 작은 burst나 누수인 경우에는 발견하는 것이 어렵다.
- 누수 발견에 사용하는 접근법은 크게 “모델 기반”과 “데이터 기반”으로 나눌 수 있다.
- 본 논문은 데이터 기반 프레임워크에서 유압 데이터(예: 압력, 흐름, 가속도)를 연속적으로 또는 준-연속적으로 모니터링하는 장기 모니터링 및 탐지 시스템에 초점을 맞추고 있다.
- 수중 음향 특징을 사전 처리하기 위한 주요 도구로써 특이 스펙트럼 분석(SSA)이라는 비모수적 방법을 사용한다.
- 높은 배경 잡음이 존재하는 WDN에서 누수를 탐지하는 준지도 데이터 기반 접근법을 제시한다.

2. Background on SSA

- 궤도 행렬의 특이값 스펙트럼을 이용한 시계열 신호 분석을 “SSA”라고 한다.
- SSA는 비모수적이며, 임베딩 차원을 나타내는 파라미터 하나만 가진다.
 - 모수를 함수형태의 분포를 가정하여 접근하는 방식이 모수적 방식
 - 반면 모집단의 분포를 가정하지 않고 접근하는 방법이 비모수적 방법

2.1 Algorithmic details of SSA

- SSA는 시계열 $\{s_n, n = 1 \dots N\}$ 의 유한 길이 레코드를 고려하고, 물리적으로 해석이 가능한 구성요소의 합으로써 s 를 분해하는 것이 목표이다.
- 이 작업은 “분해”와 “재조합”으로 이루어 진다.

2.1.1 Decomposition

- 임베딩.
- 1차원 N-샘플 시계열을 $(L, K=N-L+1)$ 의 꺾적 행렬로 만든다. 여기서 L 은 임베딩 차원이다. L 의 범위는 $[2, N-1]$ 이다.

$(L=N, K=N-L+1)$ 인 경우

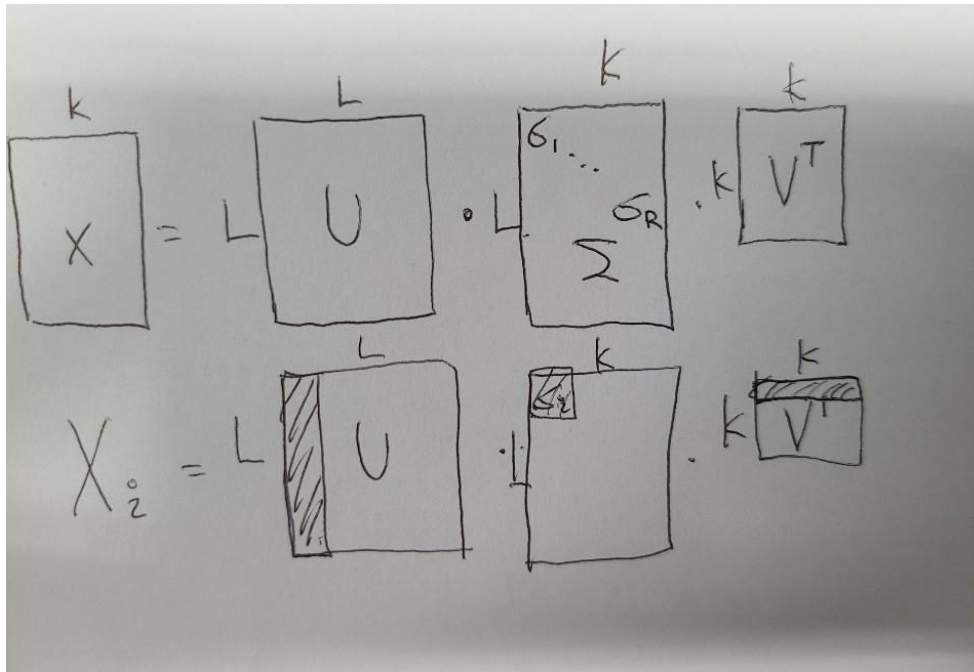
$$X = \begin{pmatrix} s_1 & s_2 & \dots & s_K \\ s_2 & s_3 & \dots & s_{K+1} \\ \vdots & \vdots & \vdots & \vdots \\ s_L & s_{L+1} & \dots & s_N \end{pmatrix} \quad (L=1, K=N-L+1) \text{인 경우} \quad X = (s_1 \ s_2 \ \dots \ s_N) \quad X = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{pmatrix}$$

- L 은 분해에 직접적인 영향을 주기 때문에 신중하게 선택해야 한다.

2.1.1 Decomposition

- Singular value decomposition(SVD)
- X 의 SVD(Singular Value Decomposition)은 아래와 같다.

$$X = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^R X_i \quad \text{with } X_i = \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$



2.1.1 Decomposition

- X 의 SVD는 lag-covariance matrix $C = XX^T$ 의 고유값 분해를 통해 얻을 수 있다.
- $C = U \Lambda U^T$ 으로 표현할 수 있다.
- Λ 는 고유값의 대각행렬이다. i 번째의 고유값은 σ_i^2 와 같다.
- 또한 $v_i = X^T u_i / \sigma_i$ 으로 V 를 추론할 수 있다.

2.1.2 Reconstruction

- Grouping.
- 기본 행렬 X 를 X ($i = 1, \dots, R$) 집합으로 분리를 한다.
- 즉, $X = \sum_{i=1}^R X_i$
- 이후, $X = \sum_{k=1}^r X_{l_k}$, ($X_{l_k} = \sum_{i \in l_k} X_i$)으로 그룹으로 묶는다.
- 즉, R 개 집합을 r 개 집합으로 묶는다.

2.1.2 Reconstruction

- Averaging.
- 이전 단계에서 그룹화된 것을 길이가 N인 새로운 시계열로 변환하는 것이다.
- $X_{l_k} = (x_{i,j})$ 의 교차 대각선 평균은 시계열 $\{x_n^{(k)}, n = 1 \dots N\}$ 의 요소를 제공한다.

$$x_n^{(k)} = \begin{cases} \frac{1}{n} \sum_{m=1}^n x_{m,n-m+1} & \text{for } 1 \leq n < L \\ \frac{1}{L} \sum_{m=1}^L x_{m,n-m+1} & \text{for } L \leq n \leq K \\ \frac{1}{N-n+1} \sum_{m=n-K+1}^L x_{m,n-m+1} & \text{for } K+1 \leq n \leq N. \end{cases}$$

- 결과 시계열을 기본 구성요소라고 한다.

2.2 Bhattacharya distance

- 도메인 X 에 걸쳐 정의된 이산 확률 분포 p 와 q 의 경우, 아래와 같이 정의된다.

$$B_d = -\ln \left(\sum_{x \in X} \sqrt{p(x)q(x)} \right)$$

- $0 < B_d < \infty$ 이고, B_d 의 값이 커질수록, 두 분포가 서로 많이 달라지고 해당하는 클래스들이 분리된다.
- 이 측정은 히스토그램의 유사도를 정량화 하기 위해 특징 값들의 두 사례 집합의 히스토그램 사이에서 계산된다.

2.3 One-class support vector machine (OCSVM)

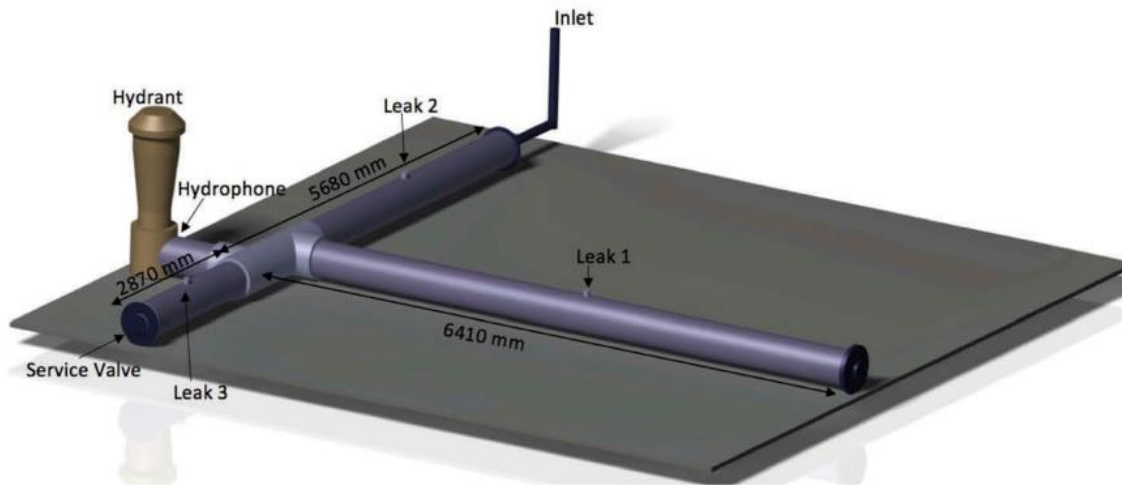
- OCSVM은 학습 시에 기본 상태의 데이터만 필요하며, 새로운 데이터는 known과 unknown으로 분류된다.
- 여기서 사용하는 많은 커널이 있지만, Gaussian radial basis functions (RBF)을 제일 많이 사용한다.

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\gamma^2}\right)$$

- γ 는 실수이고, $\|x - x'\|$ 는 새로운 데이터와 학습 데이터 간의 차이를 의미한다.
- γ 의 기본값은 $1 / \text{\#feature}$ 이다.

3.1 Laboratory testbed

- 실험실 시험대는 PVC관, T자관, 소화전으로 구성되어 있다.
- 파이프는 내경 15.24cm 이고 총 길이가 약 15m이며 3개의 누수 밸브가 있다.
- 배관의 끝 중 하나는 건물 급수 분배 시스템에 직접 연결되고, 다른 쪽은 개조된 소화전에 연결되며, 나머지는 엔드 캡으로 막혀져 있다.



3.1 Laboratory testbed

- 누수는 0.32cm 밸브를 사용하고 서비스 밸브가 있는 곳은 2.5cm를 사용하여 시뮬레이션을 진행한다.
- 모든 누수 밸브는 파이프의 중앙에 위치한다.
- 본 논문에서 사용하는 음향 데이터는 소화전 바닥에 위치한 hydrophone을 사용하여 얻었다.
- Waterloo시로부터 물의 흐름과 직접 연결된다.

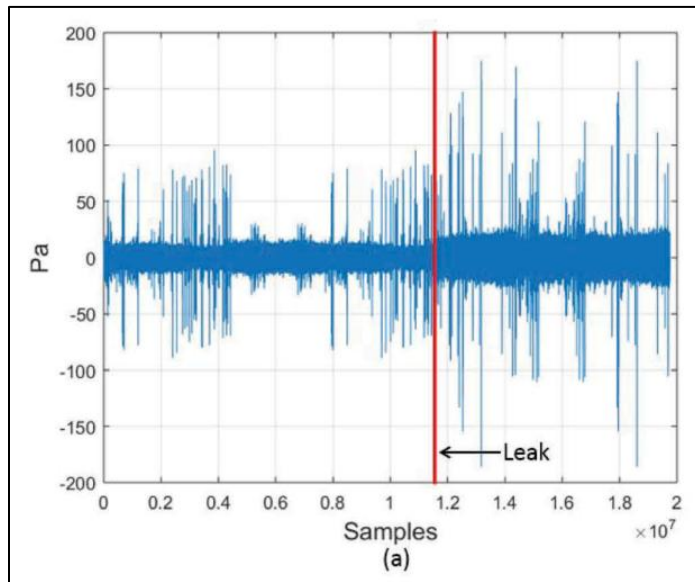


3.2 Data set description

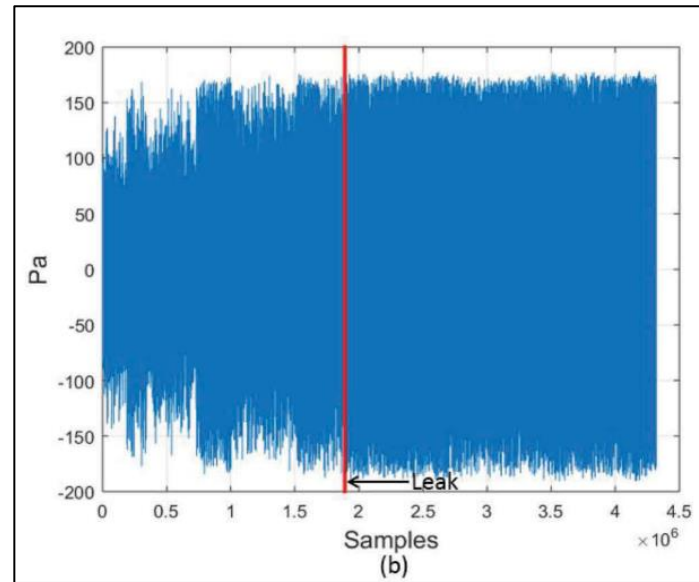
- 4가지 시나리오
 - (1) ambient(잔잔한): 누수가 없고 밸브가 닫혀 있는 경우
 - (2) 누수가 존재하며 밸브가 닫혀 있는 경우
 - (3) 밸브가 열려 있고 누수가 존재하지 않은 경우
 - (4) 밸브가 열려 있고 누수가 존재하는 경우
- 각 시나리오에서, 음향 신호는 1.35kHz의 샘플링 주파수로 얻었다.
- 그 데이터는 하루 동안 서로 다른 시간에서 10~30초 간격으로 최대 2분간 수집되고 한달동안 진행되었다.
- 이를 통해 데이터의 변동 대부분을 포착하고 특정 시간에 대한 오버피팅을 최소화한다.

3.2 Data set description

시나리오 1 & 2



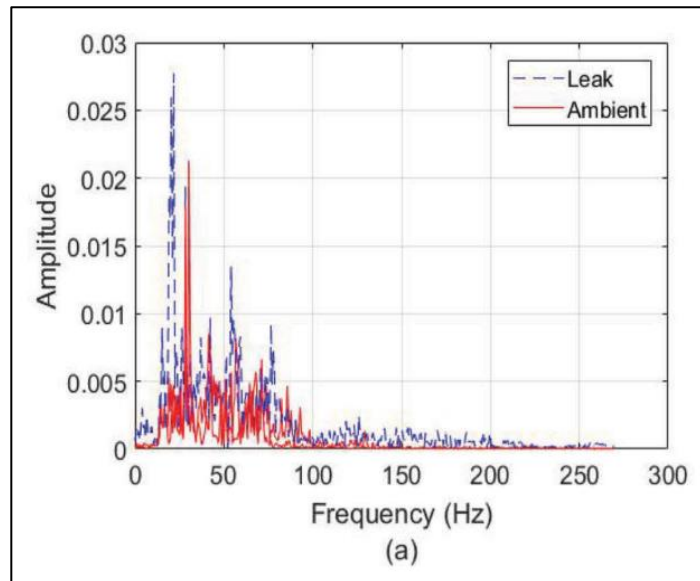
시나리오 3 & 4



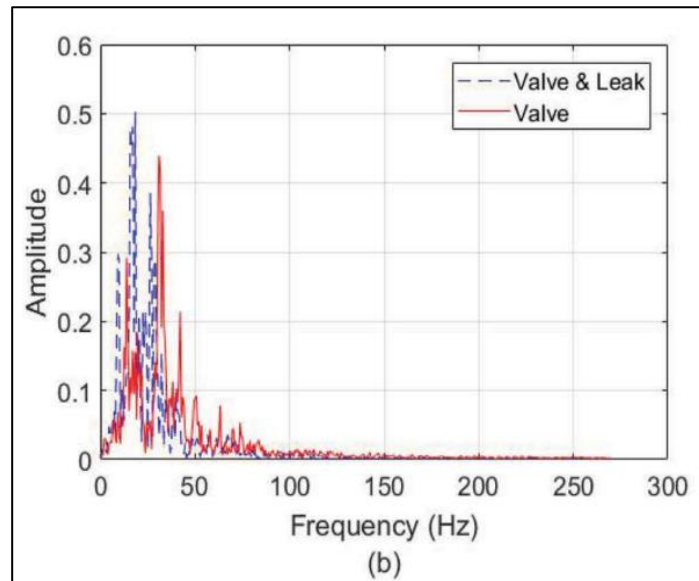
- 밸브를 열어놓은 (b)와 같은 경우는 누수를 탐지하기 어렵다.

3.2 Data set description

시나리오 1 & 2



시나리오 3 & 4

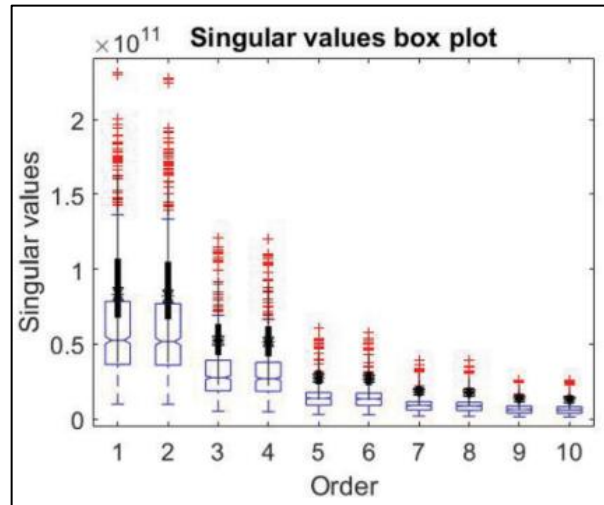


- 각각 L1, L2, L3의 누수가 발생한 경우를 하나의 데이터로 합쳤다.
- 밸브를 열면 진폭이 5~10배 상승하고 더욱 분류하기 어려워진다.

4.1 Features analysis

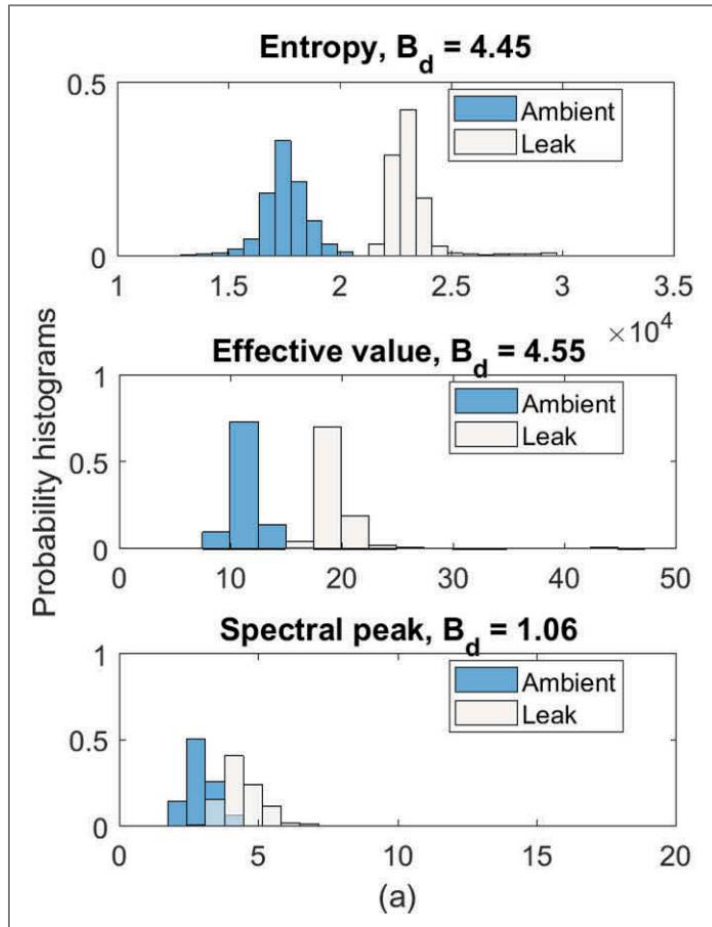
- 신호를 각각 5000개 샘플의 연속 세그먼트로 분할한다.
- 각 세그먼트에 대해 엔트로피, 유효값 및 스펙트럼 피크를 계산한다.
- 이후 히스토그램 유사도를 정량화 하기 위해, 누수 특징 값의 히스토그램과 누수가 없는 특징 값의 히스토그램을 Bhattacharya 거리를 계산한다.
- 이후 임베딩 차원 $L = 500$ 인 SSA를 각 프레임에 적용한다.
- L 은 일반적으로 경험적으로 정해지고 주어진 진동을 추출하기 위해 L 을 보통 주기보다 큰 값으로 정한다.
- Window는 전체의 신호 스펙트럼(3Hz-300Hz)가 포함되도록 크게 설정한다.
- 따라서 5000개 샘플이 500개 독립적인 기본 구성요소로 분해된다.

4.1 Features analysis



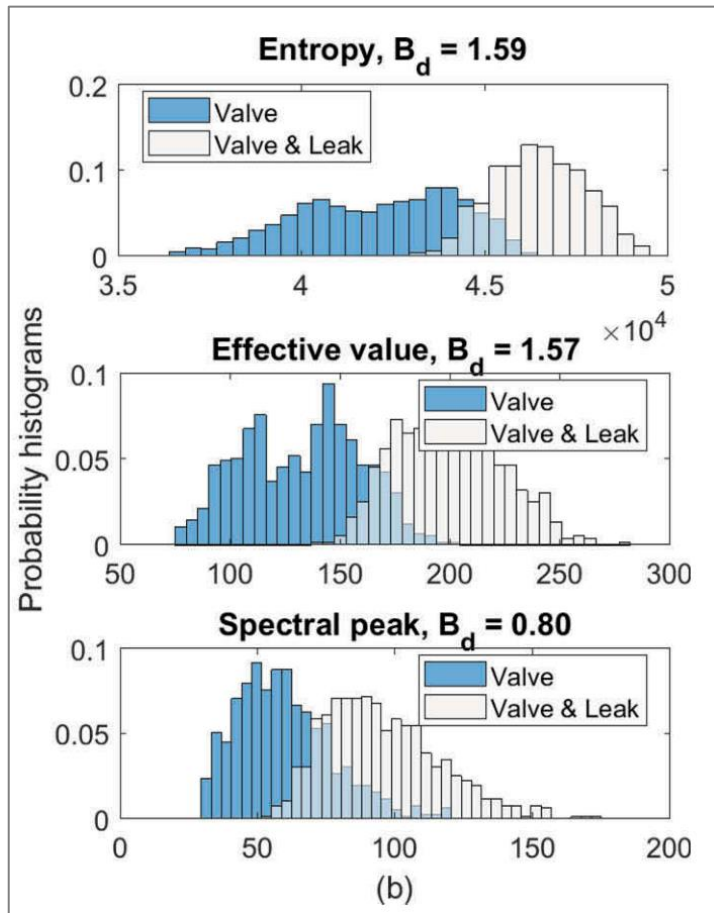
- 밸브 개방 경우에 해당하는 데이터에 대해 계산된 10개의 특이 값을 보여준다.
- 1과 2는 검은색박스(누수)와 파란색박스(정상)가 겹쳐진 부분이 많다는 것은 누수를 찾는 경우에 민감하지 않는 것을 의미한다.
- 5와 6은 검은색박스(누수)와 파란색박스(정상)가 겹쳐진 부분이 없다는 것은 누수를 찾는 경우에 민감하다는 것을 의미한다.

4.1 Features analysis



- 밸브가 닫힌 경우
- $l_1 = \{21, \dots, L\}$

4.1 Features analysis



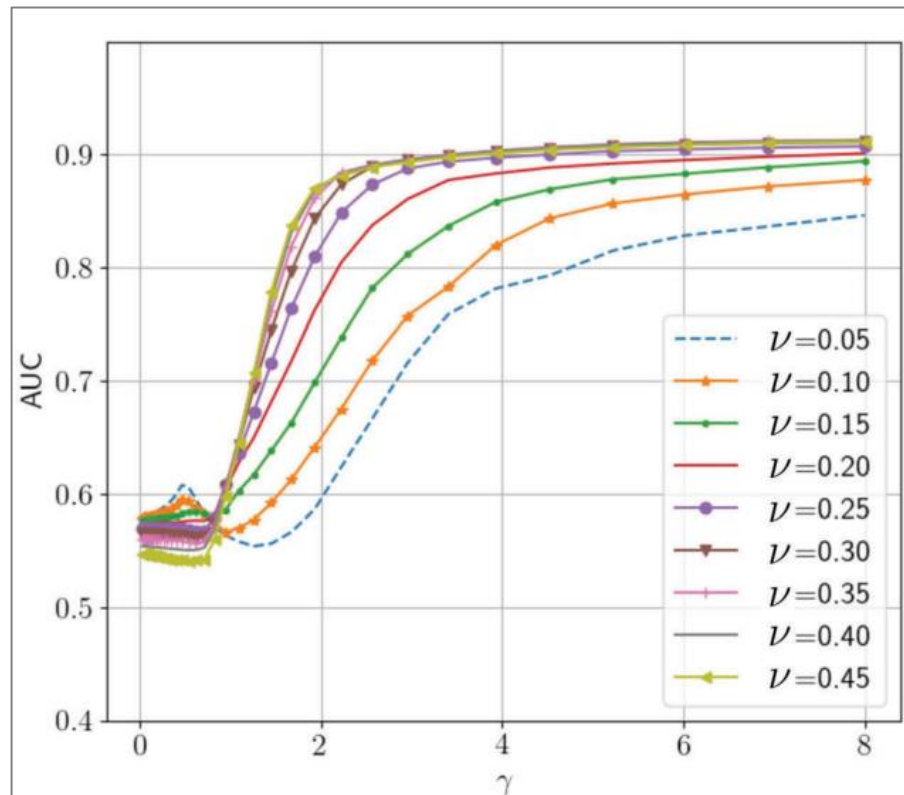
- 밸브가 열린 경우
- $l_1 = \{5, \dots, 20\}$

4.2 SSA-based classification

- (1) 누수 없는 수중 음향 신호는 프레임으로 분할되고 SSA가 적용된다.
- (2) 각 프레임은 미리 선택된 SSA 기본 구성 요소에서 재구성되는 두 가지 구성 요소를 제공하며, 하나는 그룹 l_1 에 해당하고, 두 번째는 그룹 l_2 에 해당한다.
- (3) 그런 다음 얻은 구성요소에서 특징을 계산한다. 계산된 특징 값의 30%는 무작위로 선택하여 test에 사용한다. 가우스 커널을 사용하는 OCSVM 모델은 나머지 70% 특징 값을 사용하여 train한다.
- Leak detection 정확도는 누수 및 누수 없는 데이터를 사용하여 ROC 곡선을 통해 평가
- AUC는 leak detection 정확도의 지표이다.

4.2 SSA-based classification

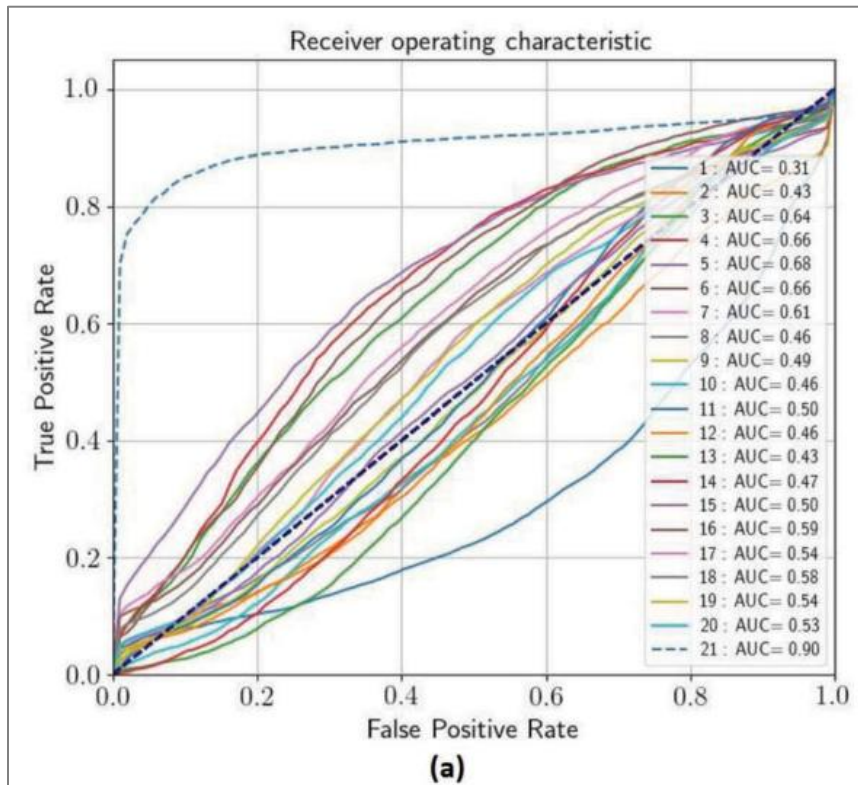
- AUC는 서로 다른 파라미터(ν, γ)으로 평가된다.
- 두 파라미터(ν, γ)는 값이 커야 좋다는 것을 알 수 있다.



4.3 SSA-based semi-supervised leak detection

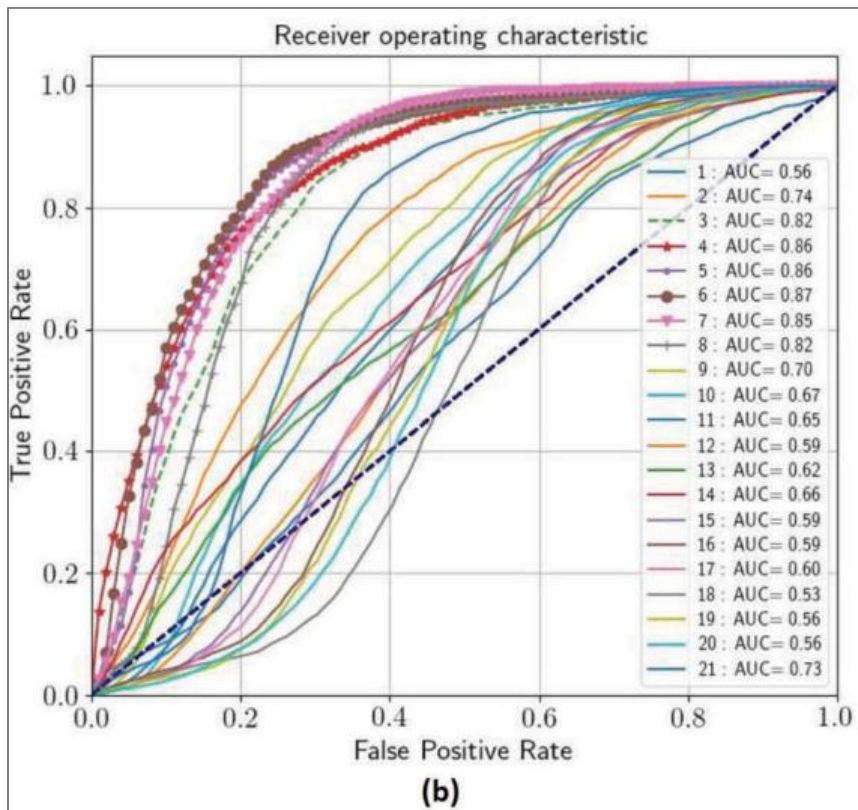
- 기본 구성 요소는 신호 요소 및 소음 요소로 분리할 수 있다.
- 신호 구성 요소는 처음의 d 개이고 나머지 $(L-d)$ 개로 소음 구성 요소를 만들 수 있다.
- 이 접근은 SSA 구성요소를 누수에 민감하고 덜 민감한 구성요소로 분리할 수 있는 가정에 근거한다.
- 실험에서는 $L = 500$, $d = 40$ 으로 하여 진행한다.
- 21개의 구성요소를 추출하고, 각 구성요소에 대해 누수가 없는 데이터를 사용하여 OCSVM을 학습한다.
- 이를 검증하기 위해 누수가 없는 데이터와 누수가 있는 데이터를 test한다.

4.3 SSA-based semi-supervised leak detection



- 밸브가 닫힌 경우
- 21번째 구성요소의 AUC 값이 높다.
- 즉, 누수에 민감한 구성요소이다.
- 반면 나머지들은 무작위 값 주위에 몰려 있다.

4.3 SSA-based semi-supervised leak detection



- 밸브가 열린 경우
- 많은 구성요소들이 AUC값이 0.6보다 작다.
- 하지만 몇몇 구성요소들(2, 9, 21)은 0.7보다 크다.
- 반면 3, 4, 5, 6, 7, 8 구성요소들은 0.8보다 크고 4, 5, 6은 0.85보다 크다.
- 따라서 누수가 발생하게 되면, 4, 5, 6 구성요소들이 의사결정을 할 것이다.

5. Conclusions

- 음향 신호에 SSA를 사용하는 연구 결과는 강한 배경 소음과 결합된 누수를 탐지하는 데 상당한 가능성을 보여준다.
- 누수에 민감한 구성 요소를 미리 선택하는 것 대신 각 SSA 구성 요소에 대해 모델을 교육하는 앙상블 모델링이 제안된다.
- 실험 환경에서 만든 데이터는 실제 데이터와 많은 차이가 있다.

Leak detection in water distribution pipes

using singular spectrum analysis

감사합니다