



Comparison of ANN model's prediction performance according to the level of data uncertainty in water distribution network

Jang, Hyewoon^a · Jung, Donghwi^{b*} · Jun, Sanghoon^c

^aUndergraduate Student, School of Civil, Environmental and Architectural Engineering, Korea University, Seoul, Korea

^bAssistant Professor, School of Civil, Environmental and Architectural Engineering, Korea University, Seoul, Korea

^cResearch Professor, Future and Fusion Lab of Civil, Environmental and Architectural Engineering, Korea University, Seoul, Korea

Paper number: 22-108

Received: 4 November 2022; Revised: 30 November 2022; Accepted: 1 December 2022

Abstract

As the role of water distribution networks (WDNs) becomes more important, identifying abnormal events (e.g., pipe burst) rapidly and accurately is required. Since existing approaches such as field equipment-based detection methods have several limitations, model-based methods (e.g., machine learning based detection model) that identify abnormal events using hydraulic simulation models have been developed. However, no previous work has examined the impact of data uncertainties on the results. Thus, this study compares the effects of measurement error-induced pressure data uncertainty in WDNs. An artificial neural network (ANN) is used to predict nodal pressures and measurement errors are generated by using cumulative density function inverse sampling method that follows Gaussian distribution. Total of nine conditions (3 input datasets × 3 output datasets) are considered in the ANN model to investigate the impact of measurement error size on the prediction results. The results have shown that higher data uncertainty decreased ANN model's prediction accuracy. Also, the measurement error of output data had more impact on the model performance than input data that for a same measurement error size on the input and output data, the prediction accuracy was 72.25% and 38.61%, respectively. Thus, to increase ANN models prediction performance, reducing the magnitude of measurement errors of the output pressure node is considered to be more important than input node.

Keywords: ANN (Artificial Neural Network), Data uncertainty, Water distribution network, Prediction

상수도관망 내 데이터 불확실성에 따른 절점 압력 예측 ANN 모델 수행 성능 비교

장혜운^a · 정동휘^{b*} · 전상훈^c

^a고려대학교 건축사회환경공학부 학사과정, ^b고려대학교 건축사회환경공학부 조교수, ^c고려대학교 미래건설환경융합연구소 연구교수

요 지

안정적인 수도 공급을 위한 상수도관망의 역할이 더욱 주목받음에 따라 비정상 상황에 대한 신속한 탐지와 적절한 대처 역시 중요시되고 있다. 장치에 의존한 탐지기법 등 기존의 방법론에는 한계가 존재하므로 데이터를 이용한 모델 기반의 방법이 개발되었다. 하지만 상수도관망 내 측정 데이터는 불확실성을 가져 실제 사용량과 다르다. 따라서 본 연구에서는 기계학습 방법의 하나인 인공신경망 모델을 이용하여 상수도관망 압력값을 예측함에 있어 데이터 불확실성의 영향을 조사한다. 정규분포를 따르는 임의의 값을 고려하여 데이터에 측정치 오류를 형성하고 측정치 오류 여부 및 종류에 따라 총 9가지 데이터를 인공신경망 모델을 통해 예측해 경향성을 비교한다. 분석을 통해 데이터 불확실성이 증가할수록 모델 성능이 감소하며, 출력데이터의 측정치 오류가 모델 성능에 미치는 정도가 더 큼을 확인하였다. 특히 입력데이터와 출력데이터의 측정 오차 크기가 동일한 경우 예측 정확도는 각각 72.25%, 38.61%로 큰 차이를 보였다. 따라서 ANN 모델 예측 성능 향상을 위해서는 입력 데이터보다 출력데이터인 주 절점의 측정 오류 크기를 줄이는 것이 중요하다.

핵심용어: ANN, 데이터 불확실성, 상수도관망, 예측

*Corresponding Author. Tel: +82-2-3290-4869

E-mail: sunnyjung625@korea.ac.kr (Jung, Donghwi)

1. 서론

국내뿐만 아니라 전 세계적으로 산업이 발전하는 등 인간적 삶의 패턴이 변화되면서 필요 수요량은 점점 증가하고 있다(Lee and Choi, 2018). 자연히 고품질의 물을 안정적으로 공급하는 상수도관망의 역할 역시 주목받고 있다. 상수도관망의 구성 요소인 저수지, 배관, 절점, 펌프 등은 모두 수리학적 상관성을 가지며 연결되어 있다. 이러한 연결성은 특정 지점에서 발생한 관 파열 등의 비정상 상황이 관망 전체에 영향을 준다는 것을 의미한다. 예를 들어 상수도관에 누수가 발생하면 관에 흐르는 유량이 증가하게 되므로 수도손실이 증가한다. 이는 수용가에 평소보다 낮은 압력을 발생시켜 용수의 사용성 저하를 유발하게 된다(Jung and An, 2018). 따라서 안정적인 수도 공급을 위해서는 비정상 상황에 대해 신속히 탐지하고 적절한 대처를 취해야 한다. 그러나 다가구에 수도를 안정적으로 공급하기 위해 관망이 점점 확장되고 복잡화되며, 대부분의 관망이 지하에 매설되어 있어, 장치에 의존하여 비정상 상황을 탐지하는 방법[예, hardware-based methods (Wan *et al.*, 2022)]의 경우 시간이 많이 소요되고 정확도가 낮다는 한계점이 존재한다.

상기한 문제를 해결하기 위해, 지난 20여 년 동안 다양한 모델 기반의 누수 탐지 방법(model-based leak detection methods)이 개발되었다(Wan *et al.*, 2022). 시계열 데이터 분석 방법(Sanneh *et al.*, 2022), 통계적 공정관리 기법(Statistical Process Control, SPC) 기반 방법(Jung and Lansey, 2014), 칼만필터(Kalman Filter) 방법(Ye and Fenner, 2011; Kim *et al.*, 2013; Jung *et al.*, 2015), 인공신경망(Artificial Neural Network, ANN) 모델(Mounce *et al.*, 2006)과 장단기 기억망(Long Short-Term Memory, LSTM) 모델(Lee and Yoo, 2021) 같은 기계학습(machine learning)을 이용한 방법으로 분류할 수 있다. 모델 기반의 방법은, 정상 상황의 상수도관망을 기준으로 예측 모델을 구축하고, 그 모델로부터 압력 및 관 유량 등 다양한 변수를 예측하며, 이를 현장에서 측정한 실측값과 비교하여 그 잔차(residual)를 모니터링하여 이상을 탐지한다. 따라서 관 파열 이상탐지의 높은 정확도를 확보하기 위해서는 먼저 예측 정확도가 높은 모델을 구축해야 한다.

최근에는 물의 공급에 직접적으로 관여하는 시설들뿐만 아니라 여러 절점, 나아가서는 주거 지역에 첨단검침인프라(advanced metering infrastructure, AMI)와 같은 스마트 미터를 설치하여 실시간으로 받아오는 데이터를 활용할 수 있는 방안을 모색하고 있다(Park *et al.*, 2020). 압력값의 경우 물 사용 패턴과 연관성이 있을 뿐만 아니라 비정상 상황에서 변화

도 두드러지게 나타난다. 모든 물 사용 지점에 스마트 미터를 설치하기에는 경제적인 문제 및 데이터 과부하 현상이 발생하기 쉬워 현실적으로 불가능하므로 대부분 일부 지점에서 측정된 데이터를 기반으로 하여 관망 전체를 분석한다. 즉, 안정적이고 신뢰성이 높은 데이터를 얻어야 관망 내에서 정확도가 높은 분석이 가능하다.

하지만 상수도관망 측정 데이터는 항상 불확실성을 갖는다. 대표적인 예로 상수도관망의 모든 계측값(예, 유량 및 압력)은 측정치 오류(measurement error, ME)를 갖기 때문에 미터기로부터 측정되는 데이터는 그의 실제 사용량과 다르다. 특히, 시공간적으로 높은 해상도를 갖는 스마트 미터 데이터의 경우 측정치 오류는 그 크기가 수요량에 반비례한다(Badger Meter, 2022). 따라서 상수도관망 데이터의 불확실성이 모델에 주는 영향을 분석하는 작업은 필수적이다.

본 연구에서는 압력 데이터의 불확실성이 인공신경망 모델의 성능에 미치는 영향성을 분석한다. EPANET 프로그램(Rossman, 2000)을 이용하여 수요량 패턴에 따라 변하는 상수도관망 내 압력 데이터를 추출하고 데이터 불확실성의 영향을 분석하기 위해 다양한 크기의 측정치 오류를 적용한 데이터를 생성한다. 이를 인공신경망 모델의 입력 데이터로 두었을 때 예측의 정확도를 비교하여 모델 성능을 평가하고자 한다.

2. 방법론

2.1 인공신경망 모델 형성

인공신경망 모델은 인간 뇌의 뉴런에서 신호를 받고 처리하는 구조를 모방하여 만든 기계학습의 한 방법이다. 뉴런은 특정 임계값을 넘어서는 자극을 받으면 신호를 전달하고 이러한 과정의 반복을 스스로 학습하여 제시된 문제를 해결한다. 이러한 자기 학습성을 모방하여 수집된 데이터를 기반으로 통계적 분류 및 회귀 분석 등을 수행하는 모델을 인공신경망이라고 한다.

인공신경망 모델은 반복적인 자기 학습을 통해 각 뉴런당 최적의 가중치를 결정한다. 적절한 가중치 갱신을 위해서는 역전파(backward propagation) 알고리즘이 사용된다. 전체 신호가 정방향(forward propagation)으로 전달되어 최종 출력값을 가지게 되면, 손실 함수(loss function)를 이용하여 오차를 계산한다. 역전파 알고리즘은 신호를 전달하는 반대 방향으로 다시 신호를 전파하며 오차를 최소화할 수 있는 가중치를 다시 찾아낸다(Fig. 1). 이러한 과정의 반복을 통해 인공신경망에 제공되는 데이터(훈련 데이터; training data)에 최

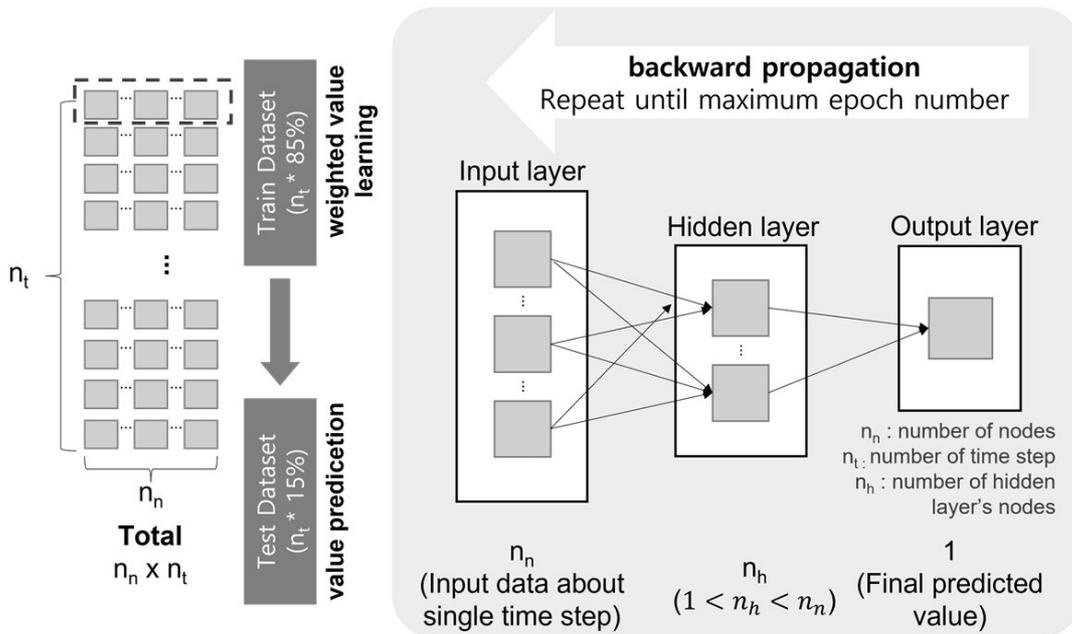


Fig. 1. Overview of ANN model

적화된 모델이 생성된다. 만약 은닉층의 레이어 수가 과도하게 증가하게 되면 최종 계산 출력값이 입력값에 미칠 수 있는 영향성이 감소하므로 역전파 알고리즘이 제대로 나타나지 않는 현상이 발생하기도 한다. 이러한 문제를 해결하기 위해서는 레이어 층 및 노드 수를 적절하게 결정하여 모델의 성능을 개선해야 한다.

인공신경망 모델의 성능을 결정하는 매개변수에는 신경망의 최소 단위인 노드(뉴런)의 수와 연결 패턴, 레이어 층의 수, 최적화를 위한 손실 함수(loss function) 등이 있다. 인공신경망 모델 안에서 입력 데이터가 입력층(input layer)으로 들어 오게 되면 가중치를 고려하여 다음 층으로 데이터를 연결한다. 최종 출력층(output layer)에서 초기에 지정해둔 손실 함수의 값이 가장 최소화될 수 있도록 학습을 통해 가중치를 갱신하며 최적의 인공신경망 모델을 구축한다. 이때, 인공신경망은 epoch라는 매개변수를 통해 학습의 과정을 반복하는 횟수를 지정한다.

손실 함수는 실제값과 인공신경망 모델의 예측값 사이의 오차를 계산하며 대표적으로 사용되는 함수로는 MSE (Mean Squared Error) 및 MAE (Mean Absolute Error)가 있다. MSE를 사용할 경우에는 예측값 y_{pred} 과 실제값 y 사이 특정 오차값에 따라 다른 크기의 가중치를 부여하므로 최대 및 최소값을 감소시키는 역할을 하며 이상치에 대한 민감도가 크다[Eq. (1)]. 이와 반대로 MAE의 경우에는 각 오차값에 동일한 가중치를 부여하므로 인공신경망 모델의 예측값이 실제값과 전반

적으로 얼마나 차이가 나는지 판별할 수 있다[Eq. (2)]. 상수도 관망에서는 전체적인 압력의 경향성보다 허용 가능한 범위를 벗어나는 압력값이 피해 정도를 결정하기 때문에 본 연구에서는 이상치를 최소화할 수 있는 MSE를 손실함수로 적용한다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - y_{pred})^2 \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - y_{pred}| \quad (2)$$

본 연구에서 상수도관망의 압력을 예측하기 위해 설계한 인공신경망 모델의 전체적인 개요도는 Fig. 1과 같다. 인공신경망 모델은 시간 t 에서의 주 절점(critical node)의 압력을 예측한다. 즉, 매 시간 간격마다 수집되는 주 절점의 압력 데이터가 본 연구에서 개발한 모델의 출력값(output)이다. 주 절점은 전체 상수도관망의 성질을 대표하는 절점으로, 수요량에 따라 압력의 변동 폭이 가장 크고 저수지와 가장 멀리 떨어져 외부 수원의 영향을 최소로 받는 절점으로 선정한다. 이는 EPANET 프로그램을 이용한 분석을 통해 결정한다.

인공신경망의 입력값(input)은 시간 t 일 때 주 절점의 압력과 상관성이 높은 절점들의 동시간대(t) 압력 데이터로 설정한다. 이때 입력데이터로 활용한 상관성 높은 절점의 개수는 변수 n_n 으로 설정한다. 또한, n_n 은 인공신경망에 고려된 훈련

(training) 및 시험(test) 데이터의 개수이며 이는 하루 동안 수집된 각 절점에서 수집된 압력 데이터의 시간 간격의 개수이다(예, 매 15분마다 데이터가 측정된다면 하루 동안의 n_t 는 96의 값을 가짐). 본 논문에서는 n_t 중 85%를 훈련을 위한 훈련 데이터셋, 15%를 인공지능망의 예측 결과 성능을 검증하기 위한 테스트 데이터셋으로 설정하였으며 훈련 및 시험을 위한 데이터는 무작위로 선정한다. 마지막으로 n_h 는 은닉층을 구성하는 노드의 개수를 의미한다. n_h 은 n_n 보다 하나 작은 값으로 설정한다.

2.2 측정치 오류 데이터 생성

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3a)$$

$$\Phi(x|\mu, \sigma^2) = \int_{-\infty}^x N(t|\mu, \sigma^2) dt \quad (3b)$$

$$\Phi^{-1}(q) = \text{inv } f\{x : \Phi(X) > q\} \quad (3c)$$

본 연구의 목적은 입력 및 출력 데이터의 불확실성 정도의 크기가 인공지능망의 예측 정확도에 미치는 영향을 분석하는 것이다. 이를 위해 본 절에서는 절점별 수집된 압력 데이터에 측정치 오류(ME)가 어떻게 고려되었는지 설명한다.

측정치 오류는 누적분포(Cumulative Density Function, CDF) inverse sampling 방법을 사용하여 확률분포를 따르는 임의의 값을 생성하였다(Jun *et al.*, 2021). 이를 EPANET으로부터 출력된 압력에 더하여 그 데이터를 훈련 및 시험 데이터로 활용한다. Jun *et al.* (2021)의 경우 오직 수요량 결측값 보완을 위해 CDF inverse sampling 기법을 적용하였으며, WDS 내의 압력데이터의 측정치 오류를 고려하기 위해 기법을 적용한 연구는 미비하다. 본 연구에서 고려한 압력값은 측정치 오류가 더해졌기 때문에 데이터의 불확실성이 포함되어 있으며 이는 실제 현장에서 수집되는 데이터와 그 특성이 비슷하다(Jun and Choi, 2022).

측정치 오류는 Eq. (3a)와 같이 평균 μ 및 분산 σ^2 을 갖는 정규분포를 따른다고 가정하며 이를 Eq. (3b)와 같이 누적 분포 형태로 나타낸다. 이를 다시 역함수의 형태로 취한 Eq. (3c)와 같이 정규분포를 따르는 역 누적분포함수를 이용하여 임의의 무작위수 q 를 추출하여 기존 데이터에 더해주는 방식을 이용하였으며, 측정치 오류를 고려하지 않은 경우를 ME_0 로 두었다. 역누적분포함수의 분산 값이 증가할수록 측정치 오류 값이 크게 나타나므로 분산 값에 따라 ME 데이터셋을 추가로 정의하여 분석을 진행한다. 측정치 오류가 단계별로 어떻게

생성되는지는 아래 3절에서 설명한다.

3. 데이터셋 구축 및 적용

본 연구에서 데이터셋의 구축 과정은 Fig. 2를 따른다. 적용하고자 하는 관망 모델 요소 및 물 소비 패턴 데이터를 입력하여 EPANET 프로그램(Rossman, 2000)을 통해 모의한다. 모의 결과 추출한 절점의 압력 데이터를 입출력 데이터로 두어 모델의 성능을 평가한다. 아래에서는 각 단계별 과정을 설명한다.

3.1 주거 지역 물 소비 패턴

시간에 따른 수요량 변화 패턴은 실제 호주 Queensland에서 수집된 스마트 미터 데이터를 참고하여 생성한다. 측정 일과 가구 수 등에 따라 약간의 차이는 존재하나 출퇴근 시간 전후로 수요량이 급격히 증가하며 2번의 최고점을 나타내는 공통적인 데이터 경향성을 보인다(Gurung *et al.*, 2015). 따라서 수요량 패턴 데이터의 일평균(daily average)이 1이 되며, 출퇴근 시간 전후로 두 번의 최고점을 나타내는 수요량 데이터(Fig. 2(a))를 적용하여 EPANET 프로그램을 이용해 절점 압력값을 추출한다. 이를 측정오차가 없는 기본 입출력 데이터로 사용한다. 이때 수요량 데이터의 시간 간격은 5분으로 하루 동안 총 288개의 값으로 구성된다($n_t=288$).

3.2 Mays Network

본 연구에서는 2003년 Ozger와 Mays에 의해 제안된 간단한 형태의 관망 모델을 적용한다(Ozger, 2003). Mays network라 부르는 관망은 21개의 배관, 13개의 절점, 그리고 2개의 저수지를 포함하는 가상의 관망이다(Fig. 2(b)). 관망 전체는 서로 연결되어 있는 loop형 구조를 가지므로 유량의 흐름은 여러 방향으로 나타날 수 있다. 각 배관에는 최소 두 개의 밸브가 있어 배관이 파손될 경우 전체 관망에 영향을 주지 않고 독립적으로 격리된다. 따라서 EPANET 프로그램을 통해 분석을 진행할 경우 배관의 개폐 여부를 변화시켜 독립적인 영향성을 파악할 수 있다.

Mays network를 대표할 수 있는 주 절점의 경우 외부 수원인 저수지와 먼 곳에 있으며 임의로 모의한 결과 압력 변화량이 가장 컸던 11번 절점(J11)으로 지정한다. Mays Network의 경우 매우 작고 단순한 형태의 상수도관망이므로 데이터 과부하 방지를 위해 주 절점과 가장 근접한 위치에서 연결된 3개의 절점만을 입력값으로 선택한다.

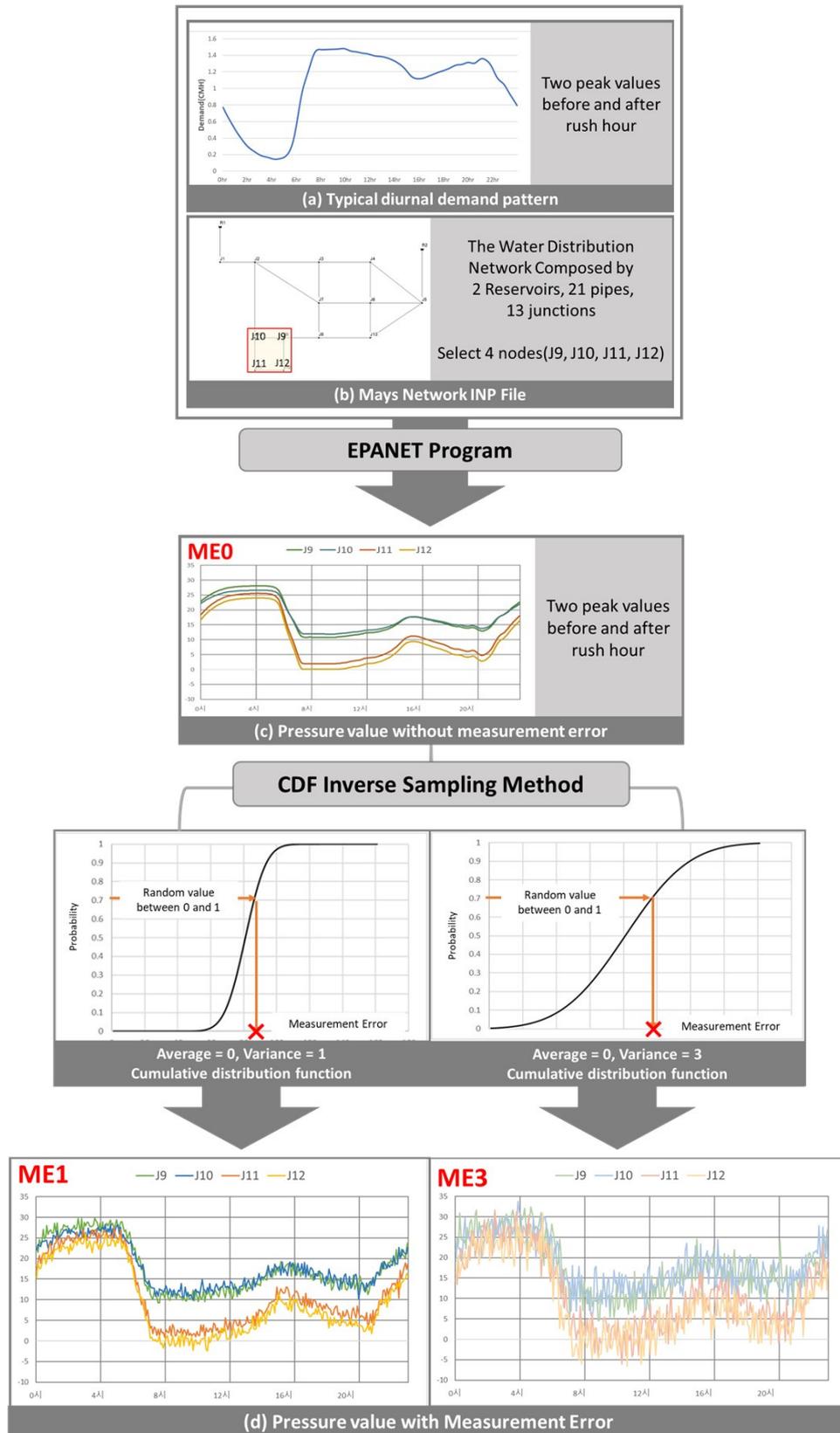


Fig. 2. Flow chart for dataset generation

3.3 데이터 분류

데이터 불확실성에 따른 수행 능력을 파악하기 위해 입력 데이터의 측정치 오류 유무와 출력 데이터의 측정치 오류 유무 및 정도에 따라 총 9가지 경우로 나누어 인공신경망 분석을 진행한다. 측정치 오류를 고려하지 않은 경우를 ME_0 , 측정치 오류의 $\mu=0$ 및 $\sigma^2=1$ 인 경우를 ME_1 , $\mu=0$ 및 $\sigma^2=3$ 인 경우를 ME_3 로 지칭한다. 각 데이터의 생성은 CDF inverse sampling 방법을 이용하여 생성하였으며, 측정치 오류의 분산이 큰 데이터를 고려한 경우가 불확실성이 더 큼을 확인할 수 있다.

이러한 입출력 데이터 조합으로 구축한 전체 데이터셋은 Fig. 3과 같다. 측정치 오류에 따라 분류한 압력 데이터 ME_0 , ME_1 , ME_3 중 입력값으로 사용한 J9, J10, J12 데이터는 ME_{in} 로, 출력값으로 사용하여 실제로 예측하고자 하는 J11 압력 데이터는 ME_{out} 의 형태로 표기하여 구분한다.

3.4 인공신경망 모델 구성

앞서 언급한 바와 같이, 3개 절점(J9, J10, J12)의 압력 데이터를 입력 데이터로 설정하였으므로 n_n 값은 3, 물 소비 패턴 데이터는 5분의 시간 간격을 갖기 때문에 인공신경망 모델의 n_t 값은 288으로 설정한다($24 - hr = 5 \text{ min} \times 288$). 추가로 은닉층의 개수는 1개이며 n_h 값은 2이다. 전체 n_t 값이 288이므로 85%인 훈련 데이터셋은 총 244개, 테스트 데이터셋은 15%인 44개의 데이터를 가진다.

그 외 최적화 방법, epoch 값, 손실 함수와 같은 모델의 성능을 결정하는 변수는 다음과 같이 설정한다. 최적화는 일반적

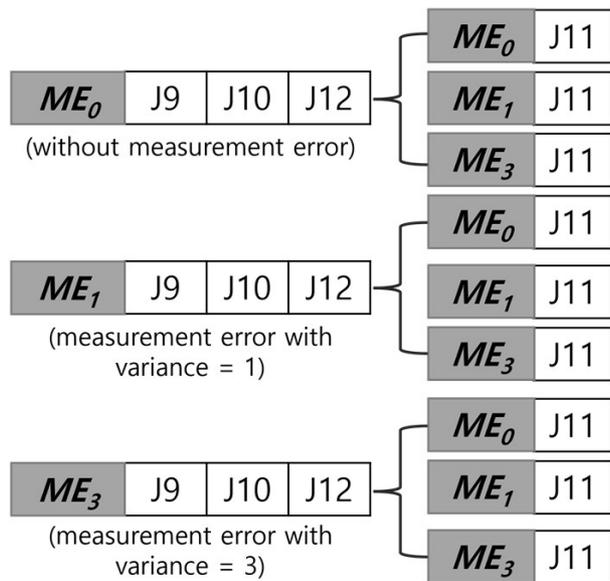


Fig. 3. Combination of total dataset

으로 가장 많이 사용하는 Adam Method를, epoch 값은 최대 1,500회를 기준으로 손실 함수의 최소값이 30회 이상 반복되면 충분히 학습되었다고 판단해 학습을 중지하도록 한다. 반복 실행 횟수에 따른 손실 함수 경향성을 그래프로 그렸을 때 epoch 1,500회 이전에서 손실 함수가 충분히 감소하여 학습 되었으므로 epoch 값을 1,500회로 설정한다.

4. 분석 결과

본 절에서는 손실 함수가 MSE인 경우, 입출력 데이터의 측정치 오류 값의 정도에 따른 인공신경망 모델의 예측 성능에 대한 결과 분석을 진행하였으며, 이를 요약하여 Table 1과 같이 정리한다. 정확도(accuracy), RMSE, 및 MAE를 모델 성능 평가를 위한 척도로 선정한다.

측정치 오류가 고려되지 않은 입력 데이터 ME_{0_in} 과 출력 데이터인 ME_{0_out} 데이터를 사용했을 때 예측 정확도가 99.51%로 나타나며 인공신경망 모델의 성능이 가장 우수하게 나타난다. 반대로 측정치 오류가 가장 큰 값일 경우 ($\sigma^2=3$) 즉, 입력 데이터 ME_{3_in} 과 출력 데이터 ME_{3_out} 데이터 조합이 예측 정확도 14.08%로 나타나며 가장 저조한 성능을 나타낸다. RMSE와 MAE는 측정치 오류가 증가할수록 증가한다. 특히, RMSE 값의 경우 입출력 데이터 모두 측정치 오류가 큰 경우에 대해 값이 급격히 증가하는 현상이 나타났는데, 이는 입출력 데이터의 예측값과 실측값 사이의 오차 최대 최소값을 최소화하는 특성이 있기 때문이라 추정한다. 따라서 압력 데이터의 불확실성이 클수록 인공신경망 모델의 예측 성능이 저하된다.

가장 좋은 성능을 보였던 ME_{0_in} 과 ME_{0_out} 데이터에 대해 인공신경망 모델을 사용하여 예측한 압력값과 EPANET 프로그램을 통해 모의한 관망 대표 절점인 J11의 압력값을

Table 1. Result of ANN model analysis according to loss function MSE

Output data		Input data	ME_{0_in}	ME_{1_in}	ME_{3_in}
ME_{0_out}	Accuracy (%)		99.51	89.25	72.25
	RMSE		0.02	0.65	2.01
	MAE		0.00	0.11	0.28
ME_{1_out}	Accuracy (%)		76.49	74.18	45.54
	RMSE		1.04	1.24	2.22
	MAE		0.24	0.26	0.54
ME_{3_out}	Accuracy (%)		38.61	32.73	14.08
	RMSE		2.99	3.00	3.95
	MAE		0.61	0.67	0.86

Fig. 4와 같이 나타냈다. 예측값과 모의한 압력값이 거의 유사하게 나타나 높은 정확도를 보임을 확인할 수 있었으나, 실제 관망 내에서 측정된 압력 데이터의 경우 외부 요인 및 측정기의 성능에 따라 측정치 오류 값이 항상 존재한다. 즉, 실제 관망

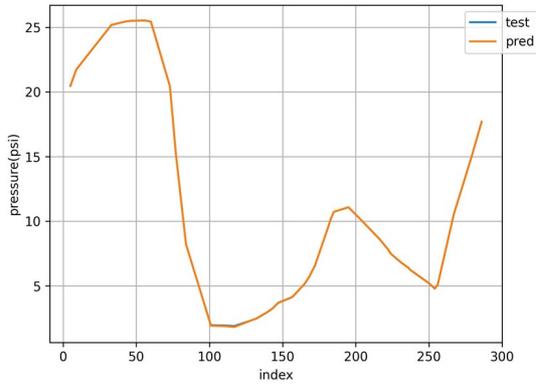


Fig. 4. Comparison between true and predicted values using ME_{0_in} and ME_{0_out}

에 대해 인공지능망 모델을 적용하는 경우 데이터 불확실성을 고려해주어야 한다.

입출력 데이터의 데이터 불확실성에 따른 경향성 파악을 위해 입력 데이터에 측정치 오류값이 고려된 경우, 출력 데이터에 측정치 오류 값이 고려된 경우로 나누어 결과 분석을 진행한다. 입력 데이터의 불확실성이 존재하는 경우에 대한 실제 압력값과 모델을 통해 예측한 값을 나타내면 Fig. 5와 같다. 입력 데이터의 측정치 오류가 작은 ME_{1_in} 데이터보다 측정치 오류가 큰 ME_{3_in} 의 경우에서 예측 압력값의 변동성이 더 크게 나타나 더 큰 이상치를 가지며, 모델의 예측 정확도 역시 89.25%에서 72.25%로 감소함을 확인할 수 있다.

출력 데이터의 불확실성 정도에 따른 성능 비교를 위해 마찬가지로 입력 경향성을 나타낸 결과는 Fig. 6과 같다. ME_{1_out} 의 경우 정확도는 76.49%, ME_{3_out} 의 경우 38.61%의 정확도를 가져, 모델 성능이 급격히 저하하여 입력 데이터와 마찬가지로 데이터 불확실성이 커질수록 압력 예측값의 이상치가 크게 나타난다.

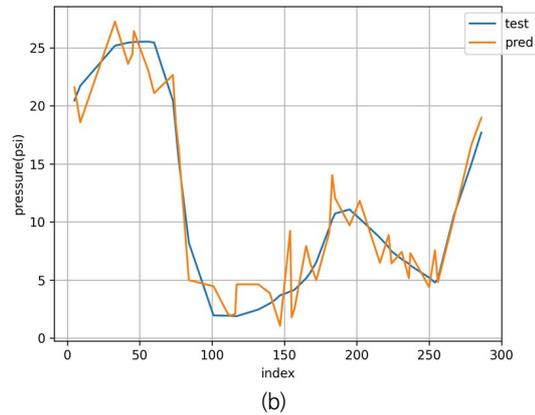
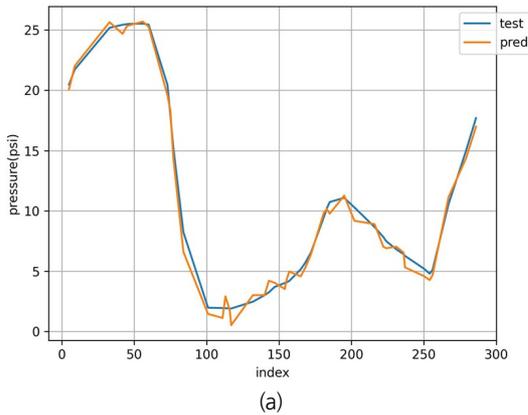


Fig. 5. Comparison between true and predicted values using (a) ME_{1_in} and (b) ME_{3_in}

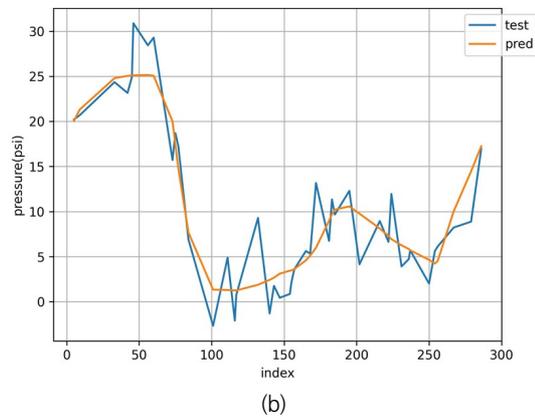
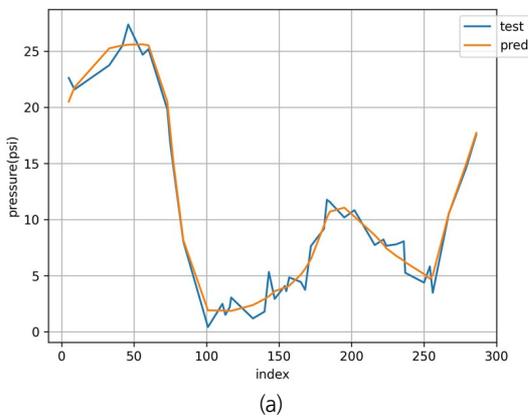


Fig. 6. Comparison between true and predicted values using (a) ME_{1_out} and (b) ME_{3_out}

모델 성능의 주 판별 요인으로 선별한 예측 정확도의 경우 Table 1에서 확인할 수 있듯, 출력 데이터에 따른 변화량이 더 크다. 따라서 입력 데이터보다 출력 데이터의 불확실성이 예측 모델 성능 저하에 더 크게 기여함을 알 수 있다. 이는 출력 데이터가 상수도관망을 대표하는 압력 데이터이므로 상수도관망 내에서 발생하는 데이터의 변동성을 잘 표현할 수 있기 때문에 출력 데이터에 따른 모델 성능의 변동성이 큰 것으로 추측한다.

5. 결론

본 연구에서는 상수도관망 내 데이터 불확실성에 따른 인공신경망 모델의 예측 성능 비교를 진행한다. 5분 간격의 1일 수요량 패턴 데이터를 통해 EPANET 프로그램을 활용하여 Mays Network의 주 절점인 J11과 가까이 위치한 상류 J9, J10, J12 압력 데이터를 활용한다. 정규분포를 따르는 임의의 값을 더해 측정치 오류를 형성한 뒤 이에 따른 인공신경망 모델의 예측 성능 변화를 파악하고자 한다.

본 연구는 상수도관망 내 데이터 불확실성 정도에 따른 인공신경망 모델의 성능 변화 및 입출력 데이터 차이에 따른 영향성 차이를 실제 모델 구축을 통해 나타냈다는 것에서 의의가 있다고 할 수 있다. 데이터의 불확실성이 커짐에 따라 모델의 예측 성능은 눈에 띄게 저하되는 것을 확인하였다. 실제 상황에서의 데이터 불확실성은 센서의 성능에 의존하므로 스마트 미터 등 압력 측정기의 성능은 관망 내 분석 결과에 큰 영향을 미칠 것이다. 또한, 전체 절점에 대한 센서 측정은 불가하므로 이를 대표할 수 있는 절점을 선정하는 것 역시 모델의 성능을 결정하는 요인으로 작용할 것이다. 특히, 입력 데이터보다 출력 데이터인 주 절점의 압력값의 불확실성 정도가 인공신경망 모델에서의 예측 정확도에 더 많은 영향을 미치기 때문에 합리적인 근거를 통해 상수도관망 내 주 절점을 선정해야 할 것이다.

향후 연구에서는 여러 종류의 상수도관망에 인공신경망 모델을 적용하여 예측 성능을 비교할 수 있다. 본 연구에서 적용한 Mays Network의 경우 단순한 형태의 관망이므로 절점 사이 압력 상관성이 높으며, 수리해석 모델로 많이 사용하는 가상의 관망이라는 한계점이 존재한다. 따라서 Mays Network보다 관망의 형태가 복잡한 실제 관망을 사용하여 분석에 사용하는 데이터를 확장하고, 데이터 규모에 따른 모델 성능의 경향성을 추가적으로 분석할 수 있다. 또한, 주 절점 및 입력 데이터 선정 방안을 체계화하여 스마트미터를 활용하는 실제 상수도관망을 대상으로 연구를 확장할 수 있다. 모의를 통한

이상적인 결과와 측정치 오류가 포함된 실제 상수도관망 데이터에 대한 모델 성능 결과를 비교하는 것 역시 흥미로운 연구가 될 것이다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2021R1A5A1032433).

Conflicts of Interest

The authors declare no conflict of interest.

References

- Badger Meter, Inc. (2022). E-series ultrasonic meters for residential applications, accessed 19 October 2022, <https://www.badgermeter.com/products/meters/ultrasonic-flow-meters/residential-ultrasonic-flow-meters/#SubNavAnchor_Documentation>.
- Gurung, T.R., Stewart, R.A., Beal, C.D., and Sharma, A.K. (2015). "Smart meter enabled water end-use demand data: Platform for the enhanced infrastructure planning of contemporary urban water supply networks." *Journal of Cleaner Production*, Vol. 87, pp. 642-654.
- Jun, S., and Choi, Y.H. (2022). "Data generation approaches to detect abnormal conditions in water distribution systems. 2." *Journal of Korea Society of Hazard Mitigation*, Vol. 22, pp. 69-79.
- Jun, S., Jung, D., and Lansey, K.E. (2021). "Comparison of imputation methods for end-user demands in water distribution systems." *Journal of Water Resources Planning and Management*, Vol. 147, No. 12, 04021080.
- Jung, D., and An, J. (2018). "Data-driven event detection method for efficient management and recovery of water distribution system man-made disasters." *Journal of Korea Water Resources Association*, Vol. 51, No. 8, pp. 703-711.
- Jung, D., and Lansey, K. (2014). "Burst detection in water system using the extended Kalman filter." *Procedia Engineering*, Vol. 70, pp. 902-906.
- Jung, D., Kang, D., Liu, J., and Lansey, K. (2015). "Improving the rapidity of responses to pipe burst in water distribution systems: A comparison of statistical process control methods." *Journal of Hydroinformatics*, Vol. 17, No. 2, pp. 307-328.
- Kim, S., Choi, D., Bae, C., and Kim, J. (2013). "Leakage detection of

- water distribution system using adaptive Kalman filter.” *Journal of Korea Water Resources Association*, Vol. 46, No. 10, pp. 969-976.
- Lee, C.W., and Yoo, D.G. (2021). “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.
- Lee, E., and Coi, B. (2018). “Transformation of paradigm of resource security through water-energy-food nexus.” *Journal of the Korean Society of Civil Engineers*, Vol. 66, No. 6, pp. 16-20.
- Mounce, S.R., and Machell, J. (2006). “Burst detection using hydraulic data from water distribution systems with artificial neural networks.” *Urban Water Journal*, Vol. 3, No. 1, pp. 21-31.
- Ozger, S.S. (2003). *A semi-pressure-driven approach to reliability assessment of water distribution networks*. Ph. D. dissertation, University of Arizona, AZ, U.S.
- Park, G.Y., Choe, J.H., and Jeong, D.H. (2020). “A study on the practical use of smart meter end-user demand data.” *Water for Future*, Vol. 53, No. 5, pp. 77-81.
- Rossman, L.A. (2000). *EPANET 2 user’s manual*. United States Environmental Protection Agency, Cincinnati, OH, U.S.
- Sanneh, J., Di Mauro, A., Venticinque, S., Santonastaso, G.F., Di Nardo, A., Wang, Y., and Lee, J. (2022). “Water demand forecasting time series data.” *Embracing analytics in the drinking water industry*, Edited by Lee, J., Keck, J., IWA Publishing, London, pp.75-122.
- Wan, X., Kuhanestani, P.K., Farmani, R., and Keedwell, E. (2022). “Literature review of data analytics for leak detection in water distribution networks: A focus on pressure and flow smart sensors.” *Journal of Water Resources Planning and Management*, Vol. 148, No. 10, 03122002.
- Ye, G., and Fenner, R.A. (2011). “Kalman filtering of hydraulic measurements for burst detection in water distribution systems.” *Journal of Pipeline Systems Engineering and Practice*, Vol. 2, No. 1, pp. 14-22.