

도시가스 배관압력 예측모델

City Gas Pipeline Pressure Prediction Model

정원희(Won Hee Chung)*, 박길주(Giljoo Park)**, 구영현(Yeong Hyeon Gu)***,
김성현(Sunghyun Kim)****, 유성준(Seong Joon Yoo)*****, 조영도(Young-do Jo)*****

초 록

도시가스 배관은 지중에 매설되어 있기 때문에 세부 관리가 어렵고 다양한 위험에 노출되어 있다. 본 연구에서는 도시가스 배관압력 실시간 데이터를 분석해 배관압력 이상을 예측하고 전문가의 의사결정을 돕는 모델을 제안한다. 국내 도시가스 공급업체들 중 하나인 중부도시가스사의 정압기에서 수집하는 실시간 배관압력 데이터와 시간변수, 외부환경변수를 통합해 분석 데이터로 사용한다. 아산시와 천안시에 위치하는 11개 정압기를 분석 대상으로 하며 분단위 배관압력 예측모델을 구현한다. Random forest, support vector regression(SVR), long-short term memory(LSTM) 알고리즘을 사용해 회귀모델을 구현한 결과 LSTM 모델에서 우수한 성능을 보인다. 아산시 배관압력 예측모델의 경우 LSTM 모델에서 RMSE가 0.011, MAPE가 0.494이며, 천안시 배관압력 예측모델의 경우 LSTM 모델에서 평균제곱근오차(root mean square error, RMSE)가 0.015, 절대평균백분율오차(mean absolute percentage error, MAPE)가 0.668로 가장 낮은 오류율을 보인다.

ABSTRACT

City gas pipelines are buried underground. Because of this, pipeline is hard to manage, and can be easily damaged. This research proposes a real time prediction system that helps experts can make decision about pressure anomalies. The gas pipeline pressure data of Jungbu City Gas Company, which is one of the domestic city gas suppliers, time variables and environment variables are analysed. In this research, regression models that predicts pipeline pressure in minutes are proposed. Random forest, support vector regression (SVR), long-short term memory (LSTM) algorithms are used to build pressure prediction models. A comparison of pressure prediction models' performances shows that the LSTM model was the best. LSTM model for Asan-si have root mean square error (RMSE) 0.011, mean absolute percentage error (MAPE) 0.494. LSTM model for Cheonan-si have RMSE 0.015, MAPE 0.668.

키워드 : 랜덤 포레스트, 배관 이상탐지, 서포트 벡터 머신

Pipeline Defects, Random Forest Regression, Support Vector Regression

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업[2014-0-00616, 대용량데이터 초고속처리 장비연구 인프라구축]의 일환으로 수행하였음.

2017년 빅데이터 플래그십 시범사업의 일환으로 수행하였음.

* Department of Computer Engineering, Sejong University(chungwh1020@gmail.com)

** Metarights Inc.(gjpark@metarights.com)

*** Department of Computer Engineering, Sejong University(yhgu@sejong.ac.kr)

**** National Information Society Agency(kimcon@nia.or.kr)

***** Corresponding Author, Department of Computer Engineering, Sejong University(sjyoo@sejong.ac.kr)

***** Korea Gas Safety Corporation(ydjo@kgs.or.kr)

Received: 2018-01-20, Review completed: 2018-03-22, Accepted: 2018-04-20

1. 서론

도시가스 배관의 안전관리는 사고예방은 물론 안정적인 공급을 위해 중요하다. 하지만 국내 33개 도시가스사가 운영하는 도시가스 배관 중 20년 이상인 노후배관이 30%에 이르며 이에 대한 관리 시스템이 미흡한 실정이다. 더불어 배관이 지중에 매설되어 있기 때문에 세부 관리가 어렵고, 굴착공사에 안전이 노출되어 있다. 현재 배관손상을 관리하기 위해 운영하는 시스템은 전기방식으로 배관피복에 일정 전위를 유지해 피복손상으로 인한 배관부식을 방지하는 방법이다. 하지만 이 시스템에는 두 가지 문제점이 있다. 첫 번째는 배관피복 손상 점검이 현장점검에 의존하고 있기 때문에 점검간격이 1년, 5년으로 길다는 것이다. 두 번째는 배관이 도로변에 매설되어 있기 때문에 현장점검원의 안전위험이 있으며 시간과 인력소요가 크다는 것이다.

본 연구에서는 현재 운영하는 전기방식 시스템의 단점을 극복하기 위해 도시가스사에서 실시간으로 수집, 관리하는 배관압력 데이터를 사용해 분 단위로 배관압력을 예측하는 모델을 제안한다. Min et al.[12]의 연구에 따르면 데이터 분석 결과는 전문가의 상황인식과 재난대응 활동의 의사결정에 긍정적인 영향을 준다. 배관압력 예측 시스템은 배관압력 수치를 제공함으로써 도시가스 배관 안전관리 전문가의 배관 이상 의사결정을 돕는다.

분석 데이터는 배관압력과 시간변수, 외부환경변수로 이루어진다. 배관압력 데이터는 정압기에서 수집하는 1차 압력과 2차 압력으로 4~6초 간격으로 실시간 수집된다. 중부도시가스의 공급권역 중 아산시와 천안시의 11개 정압기 데이터를 사용한다. 시간변수는 배관압력의 주

기성을 분석에 사용하기 위해 파생한 변수로 이동평균과 동일시점수치로 이루어져 있다. 외부환경변수는 도시가스 사용량에 영향을 주는 요인으로 기상요인, 출산율, 유가, 시간정보로 이루어져 있다[18].

배관압력 예측모델을 위해 random forest, support vector regression(SVR), long-short term memory(LSTM) 알고리즘을 사용한다. 모델의 성능은 평균제곱근오차(root mean square error, RMSE), 절대평균백분율오차(mean absolute percentage error, MAPE)를 척도로 사용해 평가한다[3, 20].

분석결과 아산시와 천안시의 분 단위 배관압력 예측모델 모두 LSTM을 사용한 모델에서 가장 우수한 성능을 보인다. 아산시 배관압력 예측모델의 경우 LSTM 모델에서 RMSE가 0.011, MAPE가 0.494이며, 천안시 배관압력 예측모델의 경우 LSTM 모델에서 RMSE가 0.015, MAPE가 0.668로 가장 낮은 오류율을 보인다.

제 2장에서는 선행연구들을 살펴보고 본 연구의 차이점을 서술한다. 제 3장에서는 도시가스 배관압력 예측시스템을 소개한다. 제 4장에서는 추가적으로 사용한 변수들을 살펴보고, 제 5장에서는 데이터 통합과 전처리 작업에 대해 서술한다. 제 6장에서는 배관압력 예측모델의 구현과 실험결과를 살펴본다. 마지막으로 제 7장에서는 본 연구를 통한 결과와 개선방향을 살펴본다.

2. 관련 연구

배관의 안전관리를 위해 다양한 배관 데이터를 분석하는 연구들이 이루어졌다. 이 장에서는 선행연구들을 소개하고 본 연구와의 차이점

과 개선점을 서술한다. 연구들은 데이터와 모델을 기준으로 구분해 볼 수 있다. 데이터는 배관이상을 탐지하기 위해 추가 운영하는 시스템에서 측정하는 배관이상 데이터와 배관운영 과정에서 생성되는 배관운영 데이터로 구분할 수 있다. 모델은 배관상태를 정상과 비정상 클래스로 나누어 제시하는 분류모델과 배관관련 수치를 제시하는 회귀모델이 있다. 선행연구와 본 연구는 <Table 1>에서 비교하고 있다.

첫 번째로 배관이상 분류모델은 정상과 비정상 클래스 정보를 제시한다. Layouni et al.[6]의 연구와 Mohamed et al.[13]의 연구는 magnetic flux leakage(MFL) 센서 데이터 셋과 artificial neural network(ANN) 알고리즘을 사용해 배관이상 분류모델을 제안한다. Akram et al.[1]의 연구와 Lee et al.[8]의 연구는 long range ultrasonic testing(LRUT) 데이터 셋과 SVM, LS-SVM 알고리즘을 각각 사용해 배관 이상 분류모델을 제안했다. Mandal et al.[11]의 연구에서는 negative pressure wave(NPW) 데이터

셋과 SVM, PSO, EPSO, ABC 알고리즘을 사용해 모델을 구성했다. Lee et al.[7]의 연구는 MFL 센서 데이터 셋과 multi-layer perceptron(MLP) 알고리즘을 사용해 배관이상 분류 모델을 구성했다.

분류모델과 다르게 배관의 운영상태에 대한 구체적인 수치를 제안함으로써 전문가의 배관 안전 관리에 도움을 줄 수 있다. 본 연구와 같이 회귀모델을 사용한 연구는 다음과 같다. Li et al.[9]의 연구는 관 음향 신호 데이터 셋과 deep neural network(DNN) 알고리즘을 사용한 회귀모델을 제안했다. Mohamed et al.[14]의 연구와 El-Abbasy et al.[2]의 연구는 각각 MFL 센서 데이터 셋과 배관 상태 데이터, 기타요인 데이터 셋에 ANN을 적용한 회귀모델을 제안했다. Nejatian et al.[17]의 연구는 배관의 유속 데이터와 LS-SVM 알고리즘을 사용, Liu et al.[10]의 연구는 가스 수요량과 수요량 관련요인 데이터와 SVM, SVR 알고리즘을 사용했다. Hwang et al.의 연구에서는 MFL 센서 데이터

<Table 1> A summary of Recent Studies

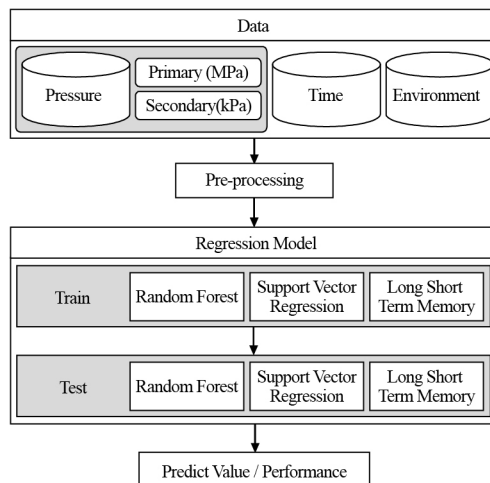
	Data		Other variables	Model	Algorithms
	Pipeline	Type			
Our Study	Management	Pressure	O	Regression	RF, SVR, LSTM
[19]		Gas demand			SVM, SVR
[17]		Condition			ANN
[18]		Velocity			LS-SVM
[15]		Sound signal			DNN
[13]	Anomaly	MFL	X	Classification	MLP
[20]					ANN
[16]					ANN
[8]					ANN
[9]					SVM, ANN
[10]		SVM			
[11]		LS-SVM			
[14]		NPW			SVM, PSO, EPSO, ABC

셋과 ANN을 사용한 회귀 모델을 제안했다[4].

본 연구에서는 배관상태에 영향을 미치는 배관압력 예측모델을 통해 전문가의 배관관리를 돕는 모델을 제안한다. Zhang et al.[21]의 연구에 따르면 배관의 내부압력은 배관의 균열, 부식에 영향을 미친다. 현재 중부도시가스사에서 정압기를 통해 실시간으로 측정, 수집하고 있는 배관의 1차 압력과 2차 압력 데이터와 배관압력 데이터의 주기성을 사용하기 위해 파생한 시간변수를 사용한다. 노승철의 연구에서 제시한 도시가스 사용량에 영향을 미치는 외부환경변수를 통합 분석해 예측모델을 구현한다[18]. 본 논문에서 제안하는 모델을 통해 전문가는 배관압력을 효율적으로 관리하고 압력이상에 대비할 수 있다. 실시간 배관압력 예측모델 구현을 위해 random forest, support vector regression(SVR), long-short term memory(LSTM) 알고리즘을 사용한다.

3. 실시간 배관압력 예측시스템

이 절에서는 실시간 배관압력 예측시스템을 소개한다. 배관압력 예측시스템 구조는 <Figure 1>과 같다. 먼저 정압기에서 수집한 배관압력 데이터와 시간변수, 외부환경변수로 이루어진 데이터 셋을 구성한다. 다음으로 통합 데이터 셋은 전처리 작업을 통해 예측모델 구현에 적합한 데이터 셋을 준비한다. 그리고 회귀모델을 구현하기 위해 세 알고리즘 random forest, SVR, LSTM을 사용한다. 마지막으로 구현한 실시간 분단위 배관압력 예측모델을 통해 예측압력을 도출하고 이를 사용해 각 모델의 성능을 RMSE, MAPE 수치로 평가한다.



<Figure 1> Pipeline Pressure Prediction System Structure

3.1 배관압력 데이터

중부도시가스사의 도시가스 배관압력 데이터는 정압기를 통해 측정한 수치로 1차 압력과 2차 압력으로 이루어져 있다. 정압기는 사용자에게 가스를 안정적으로 공급하기 위해 가스의 압력을 관리하는 시설로, 고압의 가스를 공급에 적합한 중저압으로 감압하는 시설이다. 도시가스사에서 지역 정압기로 공급하는 도시가스 배관의 압력은 1차 압력, 지역 정압기에서 일반가정으로 공급하는 압력은 2차 압력으로 구분한다.

본 연구에서는 중부도시가스사의 9개 공급권역 중 천안시와 아산시에 위치하는 지역 정압기를 분석 대상으로 한다. 천안시의 5개 정압기와 아산시의 6개 정압기의 1차 압력과 2차 압력 수치로 이루어져 있으며 5초 간격으로 측정, 수집된다.

분석을 위한 데이터 셋을 구성하기 위해 데이터 셋에 포함된 이상값과 결측값을 제거한다.

첫 번째로 수집 기간을 제한한다. 수집 초기에 발생한 이상값과 결측값을 제거하기 위해 기간은 2017년 12월 5일 00시 00분 00초부터 측정, 수집된 데이터를 사용한다. 2018년 1월 3일 00시 00분 00초까지 수집된 데이터 셋을 사용한다. 두 번째로 배관압력의 경우 1차 압력은 0.6MPa보다 큰 데이터, 2차 압력은 1.6kPa보다 큰 데이터를 분석 대상으로 한다. 결과적으로 추출한 배관압력 데이터 셋의 예시는 <Table 2>와 같다.

분 단위 예측모델을 구현하기 위해 분 단위 배관압력 데이터 셋을 구성한다. 특정 정압기의 분 압력은 해당 1분 동안 수집된 압력 수치의 평균값으로 대체한다. 각 정압기의 1차 압력 또는 2차 압력은 약 41,660건의 데이터로 이루어져 있다.

<Table 2> Pipeline Pressure Data Set Example

Time	Primary	Secondary
2017-12-05 00:00:00	0.720	2.155
2017-12-05 00:01:00	0.720	2.139
2017-12-05 00:02:00	0.721	2.131
2017-12-05 00:03:00	0.722	2.110
2017-12-05 00:04:00	0.722	2.185
2017-12-05 00:05:00	0.722	2.125
2017-12-05 00:06:00	0.723	2.123
2017-12-05 00:07:00	0.723	2.161
2017-12-05 00:08:00	0.723	2.129
2017-12-05 00:09:00	0.723	2.127
2017-12-05 00:10:00	0.723	2.147
2017-12-05 00:11:00	0.724	2.155
2017-12-05 00:12:00	0.724	2.132
2017-12-05 00:13:00	0.724	2.143
2017-12-05 00:14:00	0.725	2.138

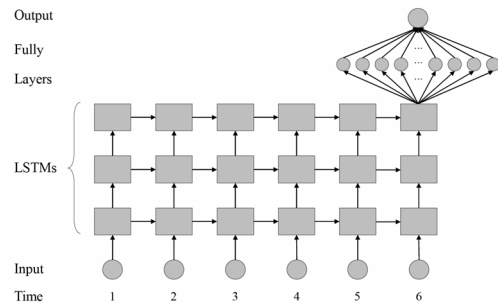
3.2 배관압력 예측모델 구현 알고리즘

배관압력 예측모델을 구현하기 위해 회귀분석에 적합한 세 가지 알고리즘 random forest, SVR, LSTM을 사용한다.

Random forest는 여러 개의 의사결정 트리를 사용하는 앙상블 알고리즘으로, 회귀모델의 경우 각 의사결정 트리에서 도출한 예측 수치들의 평균을 최종 결과로 결정한다. 앙상블 방법을 사용하기 때문에 일반화가 우수하며 노이즈에 강하다는 장점이 있다.

SVR은 support vector machine(SVM)을 회귀분석을 위해 일반화한 방법으로 초평면간의 거리에 오차 허용 범위를 적용함으로써 범위 내에서 데이터와 평면간의 거리 즉, margin을 최대로 갖는 초평면을 찾는다.

LSTM은 recurrent neural networks(RNN)의 한 종류로 시계열 데이터 분석에 적합한 인공신경망이다. RNN에서 발생하는 그라디언트 소실 문제(vanishing gradient problem)를 극복하기 위해 구조를 변형한 알고리즘으로 세 개의 게이트를 통해 상태정보를 제어한다. 본 연구에서는 3개의 층으로 이루어진 LSTM과 마지막에 fully connected network를 추가한 모델을 구현해 사용한다. 구조는 <Figure 2>와 같다.



<Figure 2> LSTM Model Structure

3.3 배관압력 예측모델 성능평가

Random forest, SVR, LSTM, 세 알고리즘에 훈련 데이터 셋과 테스트 데이터 셋을 전체 데이터 셋의 70%와 30%를 할당해 사용한다.

성능평가 척도는 RMSE와 MAPE를 사용한다. RMSE는 실제값과 예측값의 오차를 표현하는 척도로 계산방법은 식 (1)과 같다. RMSE는 데이터 수치의 크기에 영향을 받아 객관적인 평가와 현장에서 도시가스 전문가의 이해를 돕기 어렵기 때문에 본 연구에서는 MAPE를 추가로 사용한다. MAPE는 모델의 성능을 백분율로 도출해 성능을 객관적으로 평가할 수 있다. 계산방법은 식 (2)와 같다. 두 식에서 모두 y_i 는 실제 압력 수치, $\tilde{y}_i$ 은 모델이 예측한 압력 수치이다. n 은 데이터 건수를 의미한다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (1)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \quad (2)$$

4. 변수 생성 및 선정

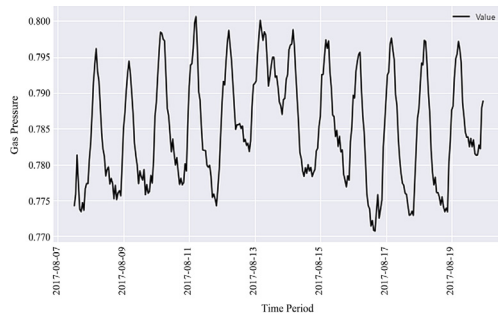
이 장에서는 배관압력 예측모델을 구현하기 위해 파생하고 수집한 변수들에 대해 서술한다. 배관압력의 주기성을 활용하기 위해 파생한 시간변수와 도시가스 사용량에 미치는 요소들인 외부환경변수로 이루어져 있다.

4.1 시간변수

배관운영 데이터는 일정 시간을 간격으로 측정된 데이터로 <Figure 3>과 같은 주기성을 가

지고 있다. 본 연구에서는 데이터의 주기성을 분석에 사용하기 위해 시계열 속성을 추가해 사용한다. 사용한 시계열 속성은 이동평균(moving average, MA)과 일정 기간 전의 수치, 최대 수치, 최소 수치이다. 이동평균 계산방법은 식 (3)과 같다. n 은 이동평균을 구할 기간, P_d 는 시점 d 에서의 수치를 의미한다.

$$MA = \frac{1}{n} \sum_{i=0}^{n-1} P_{d-i} \quad (3)$$



<Figure 3> Periodicity of Pipeline Pressure

4.2 외부환경변수

배관운영 수치는 도시가스 수요량의 영향을 받으며 수요량에 영향을 미치는 다양한 요인들이 있다. 본 논문에서는 기상정보와 유가, 출생률을 기타요인으로 선정해 분석에 사용한다.

공급권역에 대한 기상 데이터는 기상청의 기상자료개방포털에서 수집하며 일 단위 평균기온, 최고기온, 최저기온, 평균풍속, 강수량으로 이루어져 있다[16].

유가 데이터는 한국석유공사의 Petronet에서 수집하며 월 단위 데이터로 이루어져 있다[5].

출생건수는 국가통계포털에서 수집하며 월 단위 데이터로 이루어져 있다[15].

시간정보는 계절, 주말, 공휴일 여부로 이루어져 있으며, 공휴일 정보는 T developers에서 제공하는 데이터를 사용한다[19].

5. 데이터 통합 및 전처리

이 장에서는 앞에서 준비한 배관압력, 시간 변수 외부환경변수를 통합하고 전처리하는 과정을 서술한다.

5.1 데이터 통합

배관압력 데이터와 시간변수, 외부환경변수를 분석을 위해 하나의 데이터 셋으로 통합하는 작업을 진행한다. 통합은 배관압력 데이터 측정 시점을 기준으로 한다. 단, 기상요인은 일 단위, 유가와 출생률은 월 단위이므로 배관압력 데이터 단위에 맞게 각 데이터 셋을 재구성해 통합한다. 실시간 분 단위 배관압력 분석 데이터 셋은 <Table 3>과 같다.

5.2 데이터 전처리

통합 데이터 셋에 대한 정규화 작업을 진행한다. 각 변수가 분포하고 있는 값의 범위가 모

두 다르기 때문에 배관압력 수치에 주는 영향을 일정하게 조정하기 위해 정규화를 진행한다. 배관압력 수치를 제외한 독립변수들에는 최대-최소 정규화(minmax scaling)를 적용한다. 최대-최소 정규화는 최대값과 최소값을 기준으로 데이터의 분포를 조정하는 방법으로 식 (4), 식 (5)와 같다. 수식의 변수 X 는 정규화하고자 하는 수치, $X_{\min}$ 은 최소값, $X_{\max}$ 는 최대값을 의미한다. X_{std} 를 표준편차를 의미한다. 배관압력 수치를 기준으로 정규화를 진행하기 위해 최대값은 배관압력 최대값, 최소값은 배관압력 최소값으로 지정한다.

$$X_{std} = (X - X_{\min}) / (X_{\max} - X_{\min}) \quad (4)$$

$$X_{scaled} = X_{std} \times (X_{\max} - X_{\min}) + X_{\min} \quad (5)$$

다음으로 배관압력에 큰 영향을 미치는 주요 속성을 추출하기 위해 주성분 분석(PCA)을 적용한다. PCA는 데이터 분포의 분산이 큰 축을 찾는 데이터 차원축소 방법들 중 하나이다. 이를 통해 찾은 주성분들은 데이터 분포의 특징을 가장 잘 설명할 수 있는 벡터들로, 데이터 셋에서 가장 강한 특징이다. 통합 데이터 셋의 변수들 중 30개의 주성분을 추출해 배관압력 예측모델에 적용한다.

<Table 3> Variables

Type		Variables
Time		year, month, day, time, minute
Pipeline		pressure
Time	Moving average	2~5, 10, 15, 20, 25, 30min
	Same point value, Min, Max	1~5, 10, 15, 20, 25, 30min
Environment	Weather	average, min, and max temperature, win speed, precipitation
	Oil price	LPG
	Statistics	birth
	Time	weather, weekend, holiday

6. 예측모델 구현 및 실험결과

이 장에서는 앞 장에서 구성한 데이터 셋을 사용해 배관압력 예측모델을 구현한다. 회귀모델을 구현하기 위해 회귀분석 알고리즘 random forest와 SVR, LSTM을 사용한다. 아산시의 6개 정압기와 천안시의 5개 정압기에 대해 모델을 구현하고 성능을 측정한 결과는 <Table 4>와 같다.

아산시에 위치하는 6개 정압기를 대상으로 1차 압력, 2차 압력 예측모델을 구현한 결과 정압기 A의 2차 압력, 정압기 C의 2차 압력을 제외하고 모두 LSTM을 사용해 구현한 모델에서 가장 낮은 RMSE와 MAPE를 보인다. 세 알고리즘 성능의 평균을 비교한 결과 <Figure 4>와 같이 LSTM의 RMSE는 0.011, MAPE는 0.494로 가장 낮다.

<Table 5>는 각 정압기의 1차 압력, 2차 압력 예측성능이 높은 모델 추출 결과이다. 실선은 실제압력 수치, 점선은 예측압력 수치를 나타낸다. 1차 압력과 2차 압력 모두 일정한 주기성을

가지고 변화하고 있지만, 2차 압력은 수치의 변동성이 큰 것을 확인할 수 있다.

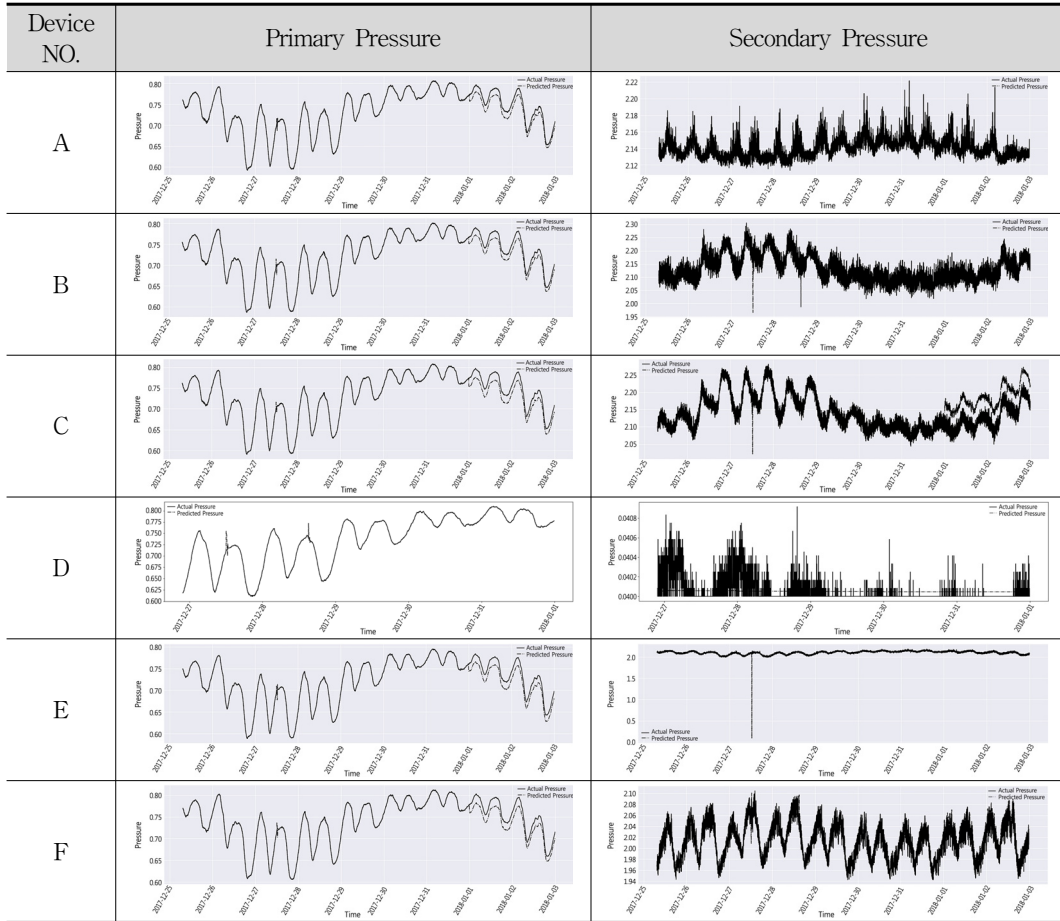
천안시에 위치하는 5개 정압기를 대상으로 1차 압력, 2차 압력 예측모델을 구현한 결과 <Table 6>과 같이 정압기 G, H, K의 2차 압력 예측모델의 경우 random forest 모델에서 낮은 RMSE와 MAPE를 보인다. 이 외의 예측모델의 경우 LSTM을 사용해 구현한 모델에서 우수한 성능을 보인다. 세 알고리즘으로 구현한 모델의 성능을 비교한 결과 <Figure 5>와 같이 LSTM으로 구현한 예측모델의 RMSE가 0.015, MAPE가 0.668로 가장 낮다.

<Table 7>은 각 정압기의 1차 압력, 2차 압력 예측성능이 높은 모델 추출 결과이다. <Table 5>와 같이 실선은 실제압력 수치, 점선은 예측압력 수치를 나타낸다. 천안시의 배관압력 수치도 아산시의 배관압력 수치와 같이 일정한 주기성을 가지고 변화한다. 하지만 1차 압력에 비해 2차 압력 수치의 변동성이 큰 것을 확인할 수 있다.

<Table 4> Asan-si Prediction Model Performances

Device NO.		Random Forest		SVR		LSTM	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
A	1	0.030	3.367	0.061	7.407	0.007	0.469
	2	0.005	0.166	0.022	0.911	0.006	0.221
B	1	0.028	3.098	0.061	7.398	0.007	0.486
	2	0.035	1.220	0.044	1.775	0.055	0.809
C	1	0.028	3.167	0.061	7.424	0.009	0.596
	2	0.026	0.865	0.059	2.532	0.031	0.984
D	1	0.046	5.112	0.058	7.009	0.002	0.088
	2	0.000	0.189	0.000	1.128	0.000	0.189
E	1	0.027	2.982	0.059	7.241	0.008	0.574
	2	0.022	0.802	0.048	2.020	0.027	0.525
F	1	0.032	3.533	0.058	7.002	0.008	0.572
	2	0.010	0.370	0.041	1.698	0.010	0.411
Avg		0.024	2.073	0.048	4.462	0.011	0.494

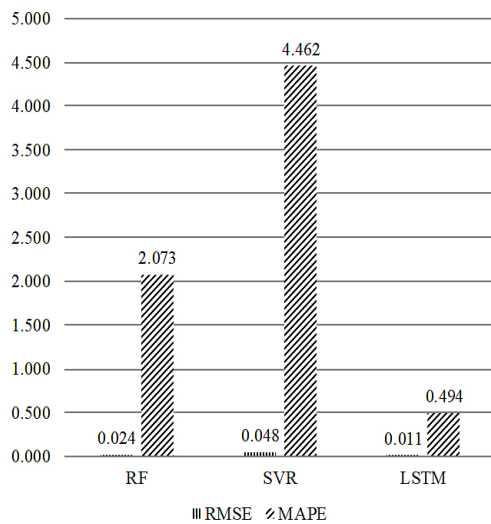
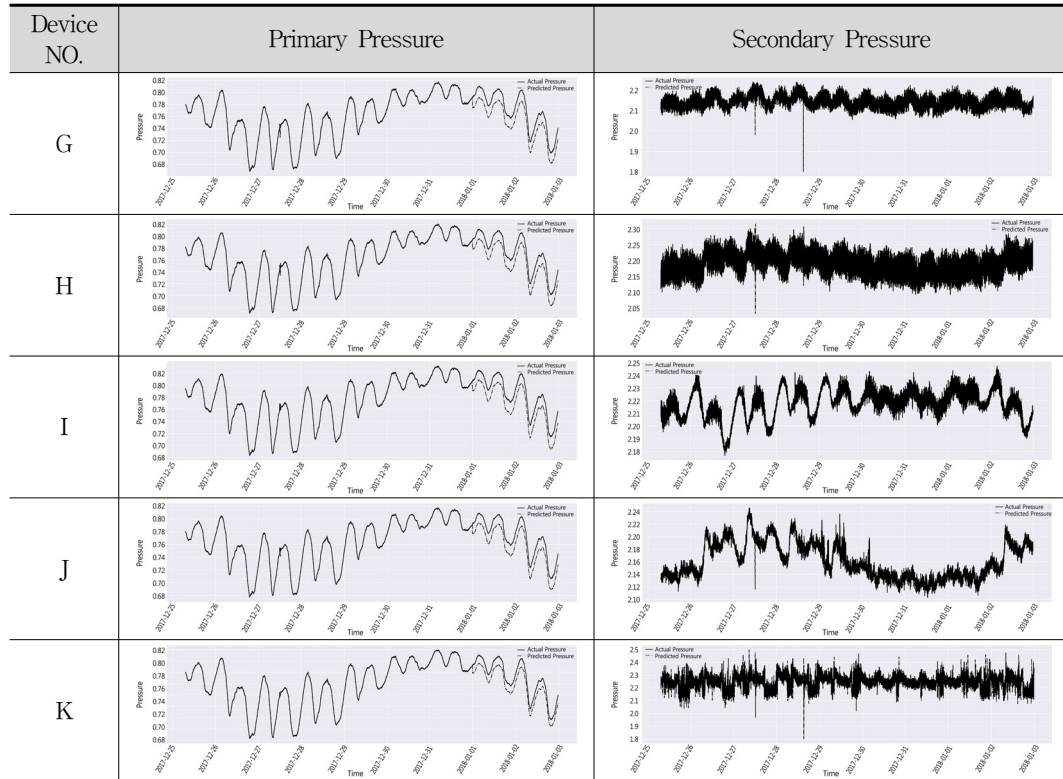
〈Table 5〉 Asan-si Pipeline Pressure Prediction



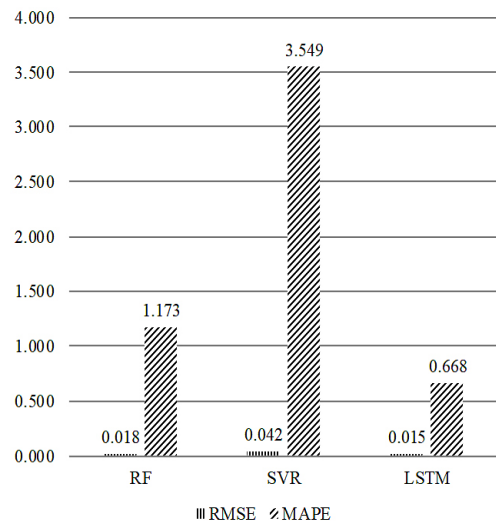
〈Table 6〉 Cheonan-si Prediction Model Performances

Device NO.		Random Forest		SVR		LSTM	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
G	1	0.018	1.763	0.053	6.031	0.009	0.596
	2	0.016	0.574	0.031	1.202	0.024	0.883
H	1	0.017	1.676	0.052	5.909	0.008	0.572
	2	0.019	0.680	0.036	1.336	0.032	1.214
I	1	0.018	1.774	0.051	5.777	0.009	0.579
	2	0.005	0.186	0.011	0.402	0.004	0.130
J	1	0.016	1.689	0.048	5.484	0.008	0.525
	2	0.027	0.883	0.043	1.692	0.006	0.214
K	1	0.017	1.692	0.053	6.095	0.007	0.453
	2	0.024	0.816	0.044	1.560	0.047	1.510
Avg		0.018	1.173	0.042	3.549	0.015	0.668

〈Table 7〉 Cheonan-si Pipeline Pressure Prediction



〈Figure 4〉 Asan-si Prediction Model Performances



〈Figure 5〉 Cheonan-si Prediction Model Performances

7. 결 론

본 연구에서는 중부도시가스사의 공급권역 중 아산시와 천안시에 위치하는 정압기를 대상으로 배관압력 예측모델을 제안한다.

분석 데이터 셋은 아산시 6개 정압기와 천안시 5개 정압기에서 5초 단위로 측정된 1차 압력과 2차 압력 데이터와 시간변수, 외부환경변수를 통합해 구성한다. 시간변수는 배관압력의 주기성을 분석에 사용하기 위해 파생한 변수로 이동평균과 동일시점의 수치, 최대값, 최소값으로 이루어진다. 외부환경변수는 도시가스 사용량에 영향을 미치는 요소들로 기상요인, 유가, 출생건수, 시간정보로 이루어져있다. 기상요인은 평균기온, 최고기온, 최저기온, 평균풍속, 강수량으로 구성되고, 시간정보는 계절, 주말, 공휴일 정보로 구성된다.

통합 데이터를 분석에 사용하기 전 전처리 과정을 거친다. 배관압력의 결측값과 이상값을 제거하고, 각 독립변수가 배관압력에 미치는 영향을 동일하게 조정하기 위해 정규화를 진행한다. 최대-최소 정규화 방법을 사용하며 배관압력의 최대값과 최소값을 범위로 한다. 그리고 PCA 방법을 사용해 배관압력에 큰 영향을 미치는 주성분들을 추출해 최종 분석 데이터 셋으로 도출한다.

배관압력 예측모델을 구현하기 위해 세 알고리즘인 random forest, SVR, LSTM을 사용하고 모델의 성능은 RMSE와 MAPE를 척도로 사용해 측정, 비교한다. 정압기 11개의 1차 압력, 2차 압력에 대한 각 예측모델을 구현하고 성능을 평가한다. 아산시의 6개 정압기 배관압력 예측 모델은 LSTM 모델의 RMSE가 0.011, MAPE가 0.494로 가장 낮아 우수한 성능을 보인다.

Random forest 모델은 RMSE가 0.024, MAPE가 2.073이고, SVR 모델은 RMSE가 0.048, MAPE가 4.462로 가장 높은 오류율을 보인다. 천안시의 5개 정압기 배관압력 예측모델 또한 LSTM 모델의 RMSE가 0.015, MAPE가 0.668로 가장 낮아 우수한 성능을 보인다. Random forest 모델은 RMSE가 0.018, MAPE가 1.173이고, SVR 모델은 RMSE가 0.042, MAPE가 3.549로 가장 높은 오류율을 보인다. 실험결과 LSTM을 사용해 구현한 예측모델의 성능이 우수함을 확인할 수 있다.

본 연구에서 제안하는 배관압력 예측모델을 통해 배관에 대한 실시간 예측압력 정보를 제공할 경우 전문가의 배관이상관리 의사결정에 도움을 줄 것으로 기대한다. 전문가는 배관압력 예측수치를 통해 배관압력의 이상을 예상하고 대비할 수 있으며, 이를 통해 배관안전 관리에 기여할 수 있다.

향후에는 정압기를 통해 측정된 배관압력과 테스트박스를 통해 측정된 배관피복의 전위를 통합 분석해 부식위치를 파악함으로써 도시가스 배관의 이상탐지에 기여할 수 있을 것으로 기대한다.

References

- [1] Akram, N. A., Isa, D., Rajkumar, R., and Lee, L. H., "Active incremental Support Vector Machine for oil and gas pipeline defects prediction system using long range ultrasonic transducers," Ultrasonics, Vol. 54, No. 6, pp. 1534-1544, 2014.

- [2] El-Abbasy, M. S., Senouci, A., Zayed, T., Mirahadi, F., and Parvizsedghy, L., "Artificial neural network models for predicting condition of offshore oil and gas pipelines," *Automation in Construction*, Vol. 45, pp. 50-65, 2014.
- [3] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," *Neural computation*, Vol. 9, No. 8, pp. 1735-1780, 1997.
- [4] Hwang, K., Mandayam, S., Udpa, S. S., Udpa, L., Lord, W., and Atzal, M., "Characterization of gas pipeline inspection signals using wavelet basis function neural networks," *NDT & E International*, Vol. 33, pp. 531-545, 2000.
- [5] Korean Oil Corporation's Petronet: <http://www.petronet.co.kr/v3/index.jsp>.
- [6] Layouni, M., Hamdi, M. S., and Tahar, S., "Detection and sizing of metal-loss defects in oil and gas pipelines using pattern-adapted wavelets and machine learning" *Applied Soft Computing*, Vol. 52, pp. 247-261, 2017.
- [7] Lee, J. Y., Afzal, M., Udpa, S., Udpa, L., and Massopust, P., "Hierarchical rule based classification of MFL signals obtained from natural gas pipeline inspection," *Neural Networks. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on IEEE.*, Vol. 5, pp. 71-76, 2000.
- [8] Lee, L. H., Rajkumar, R., Lo, L. H., Wan, C. H., and Isa, D., "Oil and gas pipeline failure prediction system using long range ultrasonic transducers and Euclidean-Support Vector Machines classification approach," *Expert Systems with Applications*, Vol. 40 No. 6, pp. 1925-1934, 2013.
- [9] Li, J., Fan, X., Chen, G., Gao, Z., Chen, M., and Li, L., "A DNN for small leakage detection of positive pressure gas pipelines in the semiconductor manufacturing," In *Online Analysis and Computing Science (ICOACS), IEEE International Conference*, IEEE, pp. 384-388, 2016.
- [10] Liu, H., Liu, D., Zheng, G., Liang, Y., and Ni, Y., "Research on natural gas load forecasting based on support vector regression," In *Intelligent Control and Automation, WCICA 2004, Fifth World Congress on IEEE*, Vol. 4, pp. 3591A-3595, 2004.
- [11] Mandal, S. K., Chan, F. T., and Tiwari, M. K., "Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM," *Expert Systems with Applications*, Vol. 39, No. 3, pp. 3071-3080, 2012.
- [12] Min, G. Y. and Jeong, D. H., "Research on assessment of impact of big data attributes to disaster response decision-making process," *Journal of Society for e-Business Studies*, Vol. 18, No. 3, pp. 17-43, 2013.
- [14] Mohamed, A., Hamdi, M. S., and Tahar, S., "A Hybrid Intelligent Approach for Metal-Loss Defect Depth Prediction in Oil and Gas Pipelines," In *Intelligent Systems*

- and Applications. Springer International Publishing, pp. 1-18, 2016.
- [13] Mohamed, A., Hamdi, M. S., and Tahar, S., "A machine learning approach for big data in oil and gas pipelines," In Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on IEEE, pp. 585-590, 2015.
- [15] National Statistics Portal: <http://kosis.kr/index/index.jsp>.
- [16] National Weather Service Station: <https://data.kma.go.kr/cmmn/main.do>.
- [17] Nejatian, I., Kanani, M., Arabloo, M., Bahadori, A., and Zendehboudi, S., "Prediction of natural gas flow through chokes using support vector machine algorithm," Journal of Natural Gas Science and Engineering, Vol. 18, pp. 155-163, 2014.
- [18] Noh, S. C., "Analysis of factors affecting energy consumption and CO₂ emissions structure in household sector," Doctoral dissertation, Seoul National University, 2013.
- [19] T developers: <https://developers.sktelecom.com/>.
- [20] Witten, I. H. and Frank, E., "Data Mining: Practical machine learning tools and techniques," Morgan Kaufmann, 2005.
- [21] Zhang, S. and Zhou, W., "System reliability of corroding pipelines considering stochastic process-based models for defect growth and internal pressure," International Journal of Pressure Vessels and Piping, Vol. 111, pp. 120-130, 2013.

저 자 소 개



정원희
2015년
2018년
2018년~현재
관심분야

(E-mail: chungwh1020@gmail.com)
세종대학교 디지털콘텐츠학과 (학사)
세종대학교 컴퓨터공학과 (석사)
세종대학교 컴퓨터공학과 (박사과정)
데이터마이닝, 기계학습, 인공지능



박길주
1999년
2011년~현재
관심분야

(E-mail: gjpark@metarights.com)
고려대학교 컴퓨터학과 (학사)
메타라이즈 주식회사 연구소장
빅데이터, 인공지능, IoT, 재난 및 안전



구영현
2004년
2006년
2006년~2014년
2014년~현재
관심분야

(E-mail: yhgu@sejong.ac.kr)
세종대학교 컴퓨터소프트웨어학과 (학사)
세종대학교 컴퓨터공학과 (석사)
세종대학교 컴퓨터공학과 (박사)
세종대학교 전임연구교수
데이터 마이닝, 오피니언 마이닝, 기계학습, 인공지능



김성현
2013년
2005년~현재
관심분야

(E-mail: kimcon@nia.or.kr)
성균관대학교 대학원 경영학과 (박사)
한국정보화진흥원 수석연구원
빅데이터, 클라우드컴퓨팅, 정보화성과관리



유성준
 1982년
 1990년
 1996년
 1998년~2000년
 2000년~2002년
 2002년~현재
 관심분야

(E-mail: sjyoo@sejong.ac.kr)
 고려대학교 전자공학과 (학사)
 고려대학교 전자공학과 (석사)
 Syracuse University 전산학과 (박사)
 한국전자통신연구원 정보검색연구팀 팀장
 ㈜서치캐스트 연구소장
 세종대학교 컴퓨터공학과 교수
 러닝 시스템, 패턴 인식, 데이터마이닝, 인공지능



조영도
 1992년
 1995년~현재
 관심분야

(E-mail: ydjo@kgs.or.kr)
 한국과학기술원 화학공학과 (박사)
 한국가스안전공사 실장
 정량적 위험평가, 위험시설 모니터링 및 위험관리,
 재난대응시스템