

도시가스 배관 위험 예측 모델 개발

A development of the gas pipeline risk prediction models

박길주* · 김영찬** · 이창열*** · 조영도**** · 정원희*****

Giljoo Park · Young-Chan Kim · ChangYeol Lee · Young-do Jo · Won Hee Chung

요 약

도시가스 배관의 안전을 위해 다양한 시스템이 가동되고 있지만 대부분 현장점검에 의존하는 한계점을 가지고 있다. 본 연구에서는 국내 도시가스 공급업체들 중 하나인 중부도시가스사의 실시간 배관운영 데이터를 분석해 배관의 위험을 예측한다. 배관의 압력, 출력전압, 출력전류, 방식전위, 전위값 데이터와 기타 도시가스 관련요인 데이터를 통합해 상관분석을 진행한다. 그리고 특정 공급권역의 실시간 배관 압력 데이터를 분석해 압력 수치를 예측한다. Random forest regression과 support vector regression(SVR) 알고리즘을 사용해 모델을 구성한 결과 배관 데이터의 시계열 정보를 추가한 데이터 셋과 random forest regression을 사용한 모델에서 가장 우수한 예측 성능을 보인다.

keywords : 도시가스 배관, 빅데이터, 위험예측

1. 서 론

도시가스 배관은 대부분 지중에 매설되어 있어 지하철, 상하수도, 전기통신, 건설 등의 여러 굴착공사에 안전이 노출되어 있다. 따라서 배관의 손상과 부식에 대한 확인이 중요한데 이 작업은 검사원의 현장작업에 의존하고 있다. 본 연구에서는 이러한 도시가스 배관안전 관리의 한계점을 극복하기 위해 배관운영 실시간 데이터를 머신러닝 알고리즘을 사용해 분석, 예측하고 전문가에게 제공함으로써 배관 이상탐지를 돕는 시스템이다.

사용 데이터는 배관운영과 관련된 데이터와 배관 상태에 영향을 주는 기타 요인 데이터이다. 배관운영 데이터는 정압기의 압력, 정류기의 출력전압, 출력전류, 방식전위 그리고 테스트박스(Test Box)의 전위 데이터 등이 있다. 그리고 기타 요인 데이터는 기상정보인 기온, 강수량, 풍속과 유가, 출생률을 사용한다(노승철, 2013). 정류기와 테스트박스 데이터를 사용한 실시간 배관 이상탐지 시스템을 통해 배관 피복의 부식을 예측한다. 정류기의 출력전압, 출력전류, 방식전위 데이터와 테스트박스의 전위값 데이터를 실시간으로 수집, 분석해 전기방식을 사용한 배관 피복의 부식을 예측한다. 더 나아가 배관의 압력과 유량을 통한 배관이상 탐지를 위해 정압기와 열량계량시스템의 데이터를 통합해 분석한다. 이 외에 배관에 영향을 미치는 배관이력, 운영환경, 외부 환경변수, 시간변수 등을 추가적으로 사용해 도시가스 배관위험 예측시스템을 구축하는 것을 목표로 한다.

* 메타라이츠 연구소장 gjpark@metarights.com

** 메타라이츠 부장 yckim@metarights.com

*** 동의대학교 컴퓨터공학과 교수 lcy@deu.ac.kr

**** 한국가스안전공사 실장 ydjo@kgs.or.kr

***** 세종대학교 컴퓨터공학과 chungwh1020@gmail.com

2. 본론

정류기와 테스트박스 데이터를 사용한 실시간 배관 이상탐지 시스템을 통해 배관 피복의 부식을 예측한다. 정류기의 출력전압, 출력전류, 방식전위 데이터와 테스트박스의 전위값 데이터를 실시간으로 수집, 분석해 전기방식을 사용한 배관 피복의 부식을 예측한다. random forest regression과 SVR을 사용한 성능은 다음 표와 같다.

표 1. 정압기 압력 예측 시스템 성능 비교

데이터셋	Random Forest Regressor		Support Vector Regressor		
	200 Trees	500 Trees	RBF Kernel	Linear Kernel	Polynomial Kernel
정압기, 기타요인	0.46	0.46	0.56	0.39	0.38
정압기, 이동평균	0.20	0.20	0.47	0.22	0.22
정압기, 이동평균, 1년전수치	0.19	0.19	0.48	0.22	0.32
정압기, 이동평균, 1년, 1년 최대, 최소	0.19	0.20	0.50	0.21	0.44

3. 결론

정압기의 압력 데이터와 기타요인 기온, 강수량, 풍속 데이터, 유가, 출생률 데이터를 통합하고 정압기 데이터의 주기성을 사용하기 위해 압력 시계열 데이터를 통합해 분석했다. 회귀분석 알고리즘은 random forest regression과 SVR을 사용해 모델을 생성, 성능을 비교했다.

실험 결과 정압기 데이터와 기타요인 데이터 그리고 압력 시계열 데이터를 모두 통합한 데이터 셋을 사용한 경우 성능이 우수했다. SVR의 경우 linear 커널함수를 사용한 모델의 RMSE가 0.21로 가장 낮았다. Random forest regression은 트리의 개수와 상관없이 SVR보다 우수한 성능을 보이며 0.19의 RMSE를 보인다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신·방송 연구개발사업 [2014-0-00616, 대용량데이터 초고속처리 장비연구 인프라구축]과 2017년 빅데이터 플래그십 시범사업의 일환으로 수행하였음.

참고문헌

노승철(2013), 가구부문의 에너지 소비요인과 이산화탄소 배출 구조 분석, 서울대학교 대학원 학위 논문